# Journal of Experimental Psychology



ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor



THE JOHN CREHAR LIBRARY
NOV 21 1952


## CONTENTS


*Learning and Integration of Component Movements in a Pattern of Motion:* G. RUBIN AND K. U. SMITH 301

*The Influence of Data Collection Procedures upon Psychophysical Measurement of Two Sensory Functions:* H. R. BLACKWELL ..... 306

*A Construction and Attempted Validation of Sensory Sweetness Scales:* S. MACLEOD ..... 316

*"Abnormal Fixation" and Learning:* H. C. WILCOXON ..... 324

*An Experimental Demonstration of Unconscious Mediated Association:* B. R. BUGELSKI AND D. P. SCHARLOCK 334

*Generalization of Extinction of an Instrumental Response to Stimuli Varying in the Size Dimension:* J. W. KLING ..... 339

*Reminiscence as a Function of the Amount of Change in the Intertrial Interval:* W. H. TEICHNER AND E. HOLDER 347

*Effects of Electroshock Convulsions on Latent Learning:* A. W. BENDIG AND R. A. PATTON 352

*The Relationship between Two Kinds of Inhibition and the Amount of Practice:* G. A. KIMBLE AND R. B. SHATEL 355

*Twenty Questions: Efficiency in Problem Solving as a Function of Size of Group:* D. W. TAYLOR AND W. L. FAUST 360



American Psychological Association

Vol. 44 No. 5 November 1952

# JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 5, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free reprints without covers. All of the cost of an author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas. Priority in publication is given to articles whose authors assume the full cost of publication.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

**Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.**

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 44, NO. 5 NOVEMBER, 1952

# LEARNING AND INTEGRATION OF COMPONENT MOVEMENTS IN A PATTERN OF MOTION[1]

GERALD RUBIN AND KARL U. SMITH

*University of Wisconsin*

Previous human engineering studies of motion have shown that the component movements in a single pattern of skilled activity are affected quite differently by the consequences of practice (**2, 4**), transfer of learning (**3**), and other conditions (**1, 2**). In the studies made, the manipulative and travel components of psychomotor activity have been segregated and automatically measured by special devices. These observations have led to the formulation of a fundamental problem of the role of learning in behavior. To what extent does learning bring about increased integration or correlation of the component movements involved in a single motion pattern?

In the experiments to be described, the correlation between component movements in a skilled task has been determined when the motion pattern is varied in complexity and when practice in the performance of different patterns is carried out over a period of several days.[2] In addition to the relation between component movements, correlations between successive days of performance and between motion patterns of different complexity have been obtained. These latter correlation values serve to point up the relative reliability of the measurements made and to give a basis of comparison for judging the effects of practice on the correlation between component movements in the motion pattern.

## METHOD

The principles and methods of dimensional and component motion analysis, as described in previous studies, have been applied to this investigation of movement integration in learning. A preplanned performance situation, designed for study of panel-control motions, has been used. This situation provides controlled variation in all critical space dimensions of the manual motion used in successive turning of switches arranged in patterns on the panel. The task for *S* consists in turning switches, 1.5 cm. in diameter, through an arc of about 45 degrees. These switches are arranged into the four different patterns of directional complexity shown in Fig. 1. The distance between switches is 16 cm. There are five manipulative movements and four travel movements in each pattern. The distance traveled in each pattern is constant. The movements are all made in the frontal plane. Separate measurement of the duration of the manipulative and travel components of motion in hundredths of a second is accomplished by means of the "Universal Motion Analyzer." This analyzer, a dual-

[1] This research was conducted with the aid of funds voted by the Legislature of the State of Wisconsin, and assigned by the Graduate School Research Committee, University of Wisconsin.

[2] The Computing Service of the University of Wisconsin provided aid and facilities to carry out the correlational analyses cited in this paper.

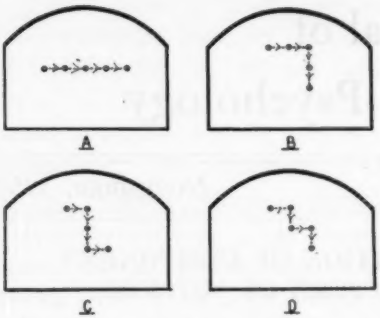


FIG. 1. The levels of directional complexity used in the experiment. Pattern A is of lowest complexity and pattern D of highest complexity. In performing these patterns of motion, *S* turns the switches on the control panel to the right. The manipulative component of motion in this experiment is the rotation of each individual switch. The travel component of motion is the transfer of the hand from one switch to another.

channel electronic relay actuated by contact with the panel switches, segregates automatically the duration of the two component movements and registers these times on two precision clocks.

In performing the task, *S* is instructed to start with the initial switch in the pattern and to turn it and succeeding switches as rapidly as possible. The duration of each component movement is recorded for the trial. Each *S* performs on all four patterns of motion on each of four successive days. Four trials on each pattern are carried out on each day. Order of running of each group of four trials on different patterns is randomized by *Ss*. An interval of about 2 min. intervened between each trial. The duration of the travel and manipulative components of movement per trial is recorded in .01 sec. Forty-eight female students of the University of Wisconsin acted as *Ss*.

The observations provided the following data. Two scores were obtained on each trial, a manipulation-time score and a travel-time score. Inasmuch as four trials were given on each pattern of motion on each day, four of each type of scores were secured on each pattern of motion in a given day's observation for every *S*. A median score was computed for the four time measures just described. All computations referred to here are based on these median scores for manipulation and travel. Four such scores, representing performance at different levels of complexity of motion, are obtained for manipulation and for travel movements on each day for each *S*. The main correlations reported here are computed by correlating these median manipulation and travel scores by *Ss*.

## RESULTS

Analysis of variance of the data on which the present study is based may be secured from the American Documentation Institute.[3] The present treatment of the data is mainly concerned with the quantitative relations between the manipulative and travel components of the motion pattern under the different conditions of the experiment. The following product-moment correlations have been computed: (*a*) between component movements as a function of days of practice, (*b*) between component movements as a function of the directional complexity of the motion pattern, (*c*) between successive days of practice for both components of movement, and (*d*) between levels of complexity of the motion pattern for both components of movement.

Table 1 indicates generally the nature of the change in the durations of the travel and manipulative motions in the task as a function of practice over the four days. The mean values given in this table represent the means for all four levels of complexity of motion and for the four trials given on each pattern of motion on each day. It will be noted that the travel movements showed about 13% reduction in duration as a function of practice, whereas the manipulative movements displayed a 24% change. The absolute variability of the travel movements does not change with practice, whereas this variability is reduced

[3] For detailed paper (or extended version of material supplementary to this article) order Document 3696 from American Documentation Institute, 1719 N. Street, N. W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 in. high on standard 35 mm. motion picture film) or $1.00 for photocopies (6 × 8 in.) readable without optical aid.

during practice over 50% in the case of the manipulative movements. In this task learning has leveled off by the third or fourth day of practice. Statistical analysis of the data, as mentioned above, indicates that the change in the time of the travel movements with practice is significant at the .05 point. The change in the durations of the manipulative movements is significant at the .01 point.

Figure 2 describes graphically the variation in level of correlation between the manipulation and travel components of motion as a function of days of practice. The parameter is level of directional complexity of the motion. The correlation between the two components of movement does not change significantly as a function of practice for any of the levels of complexity of motion. Those values marked with an asterisk indicate correlations significantly different from zero. Two such values occur on the first day of practice for the two patterns of motion of intermediate complexity. Possibly the only variation of note in these data is that all four patterns of motion give significant correlations between component movements on the fourth day of practice.

If we compare the level of correlation between component movements for all patterns of motion considered together, data like those just cited are obtained. In the order of the four

TABLE 1

DURATION OF TRAVEL AND MANIPULATIVE MOVEMENTS (IN .01 SEC.) IN RELATION TO DAYS OF PRACTICE

| Days of Practice | Travel Time | | Manipulation Time | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| 1 | .85 | .14 | 1.45 | .32 |
| 2 | .77 | .13 | 1.26 | .28 |
| 3 | .75 | .14 | 1.18 | .26 |
| 4 | .74 | .14 | 1.11 | .14 |

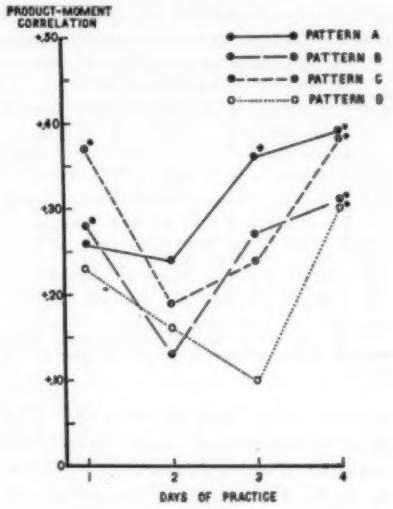


FIG. 2. Correlation between the manipulative and travel components of the same motion pattern as a function of practice for different levels of directional complexity of motion

days of practice, these values are .31, .20, .28, and .37. In obtaining these correlations, the median manipulation- and travel-time scores for each pattern of complexity are summed and mean scores for the four different days obtained. These mean manipulation- and travel-time scores for days, which combine results on all patterns of motion, are then correlated by *Ss* with one another.

The interrelation between component movements does not vary materially as a function of the directional complexity of the motion pattern. When all days of practice are considered together, the correlations between the two movements, in the order of the complexity of the motion pattern, are .35, .25, .33, and .24. In obtaining these correlations, mean manipulation- and travel-time scores for different patterns of motion complexity are computed from the scores

TABLE 2

INTERCORRELATION BETWEEN DAYS OF PRACTICE FOR THE MANIPULATIVE, TRAVEL, AND COMBINED COMPONENTS OF MOTION*

| Days | Manipulation | | | Travel | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|
| | Days | | | Days | | | Days | | |
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 2 | .78 | | | .69 | | | .76 | | |
| 3 | .70 | .89 | | .67 | .93 | | .68 | .91 | |
| 4 | .68 | .88 | .92 | .64 | .87 | .94 | .69 | .89 | .93 |

* These values are based on correlation of mean scores of all four patterns of motion for each day ($n = 48$).

over the four days of practice. The mean manipulation- and travel-time scores for different patterns of motion A, B, C, and D are then correlated with one another.

The results described are not the outcome of low day-to-day reliability of the measures of component-movement times. Values for the intercorrelations between days for the two different motion components and for both measures considered together are given in Table 2. Generally speaking, the level of correlation between successive days of practice is relatively high. Combining the time values for the two movement components does not materially raise the correlation value between days.

The low level of correlation between component parts of the same motion pattern is to be contrasted to the close relation existing between different levels of directional complexity of the motion. Table 3 gives these values of intercorrelation for different complexities of motion. These values typically exceed .90 for the two movement components and for the two parts of the motion pattern considered together. As far as the present experiment has gone, the results show that variations of motion within a given space dimension, in this case, the directional dimension, are all highly related with one another.

## DISCUSSION

The most general question considered in the experiment is that of the role of learning in determining the degree of integration between component movements required to perform a given pattern of manual motion. Results show that initially in learning, a low positive correlation of .31 exists between the manipulative and travel components of the motion pattern. This correlation value is increased slightly to .37 as a result of four days of training, at which point performance has leveled off.

The level of correlation between component movements of manipulation and travel is neither increased nor decreased systematically by varying the directional complexity of a single motion pattern.

In comparison to the specificity found for the reactive components of movement in the learning of organized motions, controlled patterns of motion of different directional complexity are highly correlated with one another. Furthermore, the techniques of control and measurement of human mo-

TABLE 3

INTERCORRELATIONS BETWEEN LEVELS OF COMPLEXITY OF THE MOTION PATTERN FOR THE MANIPULATIVE, TRAVEL, AND COMBINED COMPONENTS OF MOVEMENTS*

| Pattern | Manipulation | | | Travel | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|
| | Pattern | | | Pattern | | | Pattern | | |
| | A | B | C | A | B | C | A | B | C |
| B | .91 | | | .92 | | | .92 | | |
| C | .94 | .94 | | .94 | .92 | | .94 | .94 | |
| D | .93 | .89 | .92 | .93 | .92 | .95 | .93 | .89 | .93 |

* These values are based on correlation of mean scores of all days of practice for each pattern of motion ($n = 48$).

tion as designed for dimensional and component analysis of such motion give day-to-day reliability of measurement in limited trials of a very high level.

This research is concerned theoretically with a long-standing problem of the role of learning in defining the organization of movement in skilled motion. The findings presented show that learning has no decisive role in altering the interrelations of movements in a pattern of skilled motion. During practice component movements become neither more specific relative to one another nor more closely correlated with one another. The main changes in movement accompanying the learning in a skilled task are the changes observed in the manipulative components of the task. Changes of equal magnitude do not occur in the travel components of this task. These demonstrations of the divergent effects of learning on different component movements in the same task and of the functional insignificance of learning in defining the correlations between movements points up the need for re-evaluation of all systematic concepts of learning as they are applied to a descriptive or theoretical account of motor coordination.

## Summary

The problem of the role of learning in defining integration of separate movements in a pattern of skilled motion has been investigated by methods of dimensional analysis of motion. In the study, the correlation between the durations of the travel and manipulative components of motion in a panel-control task were determined as a function of days of practice. Results indicate that practice and learning do not alter systematically the relations between these two component movements in the task. Initially in learning, the correlation between the durations of these two component movements is low and positive. During learning over four days, this level of correlation shows no significant change.

(Received March 10, 1952)

## References

1. Davis, R., Wehrkamp, R., & Smith, K. U. Dimensional analysis of motion: I. Effects of laterality and movement direction. *J. appl. Psychol.*, 1951, **35**, 363–366.
2. Rubin, G., von Trebra, P. A., & Smith, K. U. Dimensional analysis of motion: III. Complexity of the movement pattern. *J. appl. Psychol.*, 1952, **36**, 272–276.
3. von Trebra, P. A., & Smith, K. U. Dimensional analysis of motion: IV. Transfer effects and direction of movement. *J. appl. Psychol.*, 1952, **36**, 348–353.
4. Wehrkamp, R., & Smith, K. U. Dimensional analysis of motion: II. Travel distance effects. *J. appl. Psychol.*, 1952, **36**, 201–206.

# THE INFLUENCE OF DATA COLLECTION PROCEDURES UPON PSYCHOPHYSICAL MEASUREMENT OF TWO SENSORY FUNCTIONS[1]

H. RICHARD BLACKWELL

*Vision Research Laboratory, University of Michigan*

In experiments reported elsewhere (**2**, Section I), it was demonstrated that differences in the procedure employed to collect threshold psychophysical measurements result in significant differences in the data obtained. The differences in data collection procedure employed were the response used to indicate discrimination and the order and spacing of the stimulus magnitudes utilized in the psychophysical series. Two indicator responses were employed, phenomenal report and forced choice. Phenomenal report requires *S* to make a direct evaluation of the presence or absence of discrimination. Forced choice requires *S* to indicate the presence or absence of discrimination by identifying some verifiable aspect of the stimulus other than its magnitude, and to record a guess as to the correct answer for each stimulus presentation. The order of stimulus magnitudes was varied in a number of ways and stimulus spacing was varied from a series in which stimulus magnitudes were distributed symmetrically about the threshold to a series in which magnitudes were virtually all smaller than the threshold. It was found that both the thresholds and the slopes of curves fitted to the psychophysical data varied significantly with one or another of these changes in data collection procedure.

With few exceptions, the absolute values of threshold measurements are not important. Furthermore, only the threshold values and not the slopes are usually of interest. Usually, the variation in threshold values as a function of one or another stimulus variable is of primary concern as a means of studying a sensory function. The results of the earlier experiments lead us to inquire to what extent the differences in threshold values which correspond to differences in data collection procedures are constant or to what extent they depend upon variations in one or another stimulus condition. If the differences between thresholds which depend upon procedural differences vary as a function of stimulus variables, we will obtain a different functional relation, depending upon which data collection procedure is utilized.

The present study was designed to ascertain to what extent differences between thresholds obtained with two psychophysical procedures depend upon each of two stimulus variables. The *S*s were required to discriminate brightness differences with the stimulus size and the brightness of the background field varied among experimental sessions.

Two variants of the method of constant stimuli were selected which differed in only one respect, the indicator response employed. In each case, eight stimulus magnitudes were employed in each experimental session in grouped sequences. Twenty presentations were made at first one stimulus magnitude and then another. In one case, *S* reported "Yes" when he detected a brightness difference and

[1] This research was supported by Project NR142-106, Contract N5ori-116, Task Order V, between the University of Michigan and the Office of Naval Research, U. S. Navy.

"No" when he did not. In the other, *S* had to indicate detection of a brightness difference by correctly identifying the time interval, out of four possible intervals, in which the difference occurred. The results of the earlier experiment suggested that the forced-choice procedure selected had greater session-to-session reliability and greater freedom from the influence of irrelevant variables than did the "Yes-No" procedure. In fact, the "Yes-No" procedure was shown to permit two invalid modes of response. It was believed that differences in the thresholds obtained with these methods might be dependent upon stimulus conditions to a maximum extent because of the special opportunities for the operation of invalid modes of response offered to *S* by the "Yes-No" method. Thus, the present study was intended to provide reasonable opportunity for the differences of interest to manifest themselves.

Four stimulus sizes were studied at each of two brightnesses of the background field with each of the psychophysical procedures. Comparisons were made between the thresholds obtained with each procedure at each of the eight stimulus conditions.

## Method

*Apparatus.*—A large experimental room was utilized which was equipped with opaque window shades to exclude daylight. The *S*s were seated in four theatre chairs at one end of the room. At the other end of the room a square white screen was mounted, constructed by spray-painting a galvanized iron sheet with numerous layers of white sphere paint. The average distance from *S*s' eyes to the screen was 25.6 ft. At this distance, the screen subtended 7.38° on a side. The center of the screen was pierced by a .5-in. hole. Behind the hole, a lighting system was arranged to provide an illuminated fixation and orientation spot. The lighting system consisted of a 100-w., 110-v. projection lamp, a red plastic filter, and a flashed opal glass. The lamp was in series with a system of stepped and variable resistances so that the light output was variable over an exceedingly wide range. The red filter was the standard Navy plastic utilized in dark adaptation goggles, manufactured by the Polaroid Corporation. Thus, the system provided a small red fixation spot of variable intensity.

The screen could be illuminated by each of two projection systems which were located in separate anterooms located behind *S*. The background projector was a standard Bausch and Lomb Balopticon, with a square aperture of such a size that the projected beam just covered the white screen. The stimulus projector was based upon a Balopticon, modified to permit intensity reduction and controlled exposure duration. This projector presented circular stimuli at a fixed position to the right of the fixation spot. Four apertures were employed in the spot projector in various experiments, which provided images subtending 4.20, 10.5, 28.0, and 77.5 min. of arc at *S*s' eyes. The stimulus centers were in each case 2.64° from the center of the fixation spot.

Both projectors were standardized for 2360° K. color temperature from a standard supplied by the Electrical Testing Laboratories of New York.

In order to vary the stimulus intensity to a value approximating the threshold, Wratten neutral gelatine filters mounted in glass were placed in a holder near the plane of the projection lens of the stimulus projector. In order to vary the brightness within a comparatively small range near the threshold, a filter selector consisting of two rotatable wheels was employed. One wheel was pierced by six apertures, five of which accommodated Wratten neutral gelatine filters of different low densities. The sixth "aperture" was closed by an opaque plate to permit the presentation of catch stimuli. The second wheel was pierced by eight apertures, each of which accommodated clear glass plates, the number varying among the apertures from one to eight plates. By combining a Wratten filter from one wheel with a given number of clear plates from the second wheel, it was possible to reduce the stimulus brightness in steps of approximately 8% over a total range of ten to one. It was thus possible for stimulus magnitudes to be selected to provide an adequate set of psychophysical data under all stimulus conditions.

The aperture, whose projected image was the stimulus, was mounted forward of all the filtering materials so that scattered light within the filters and glass plates would not blur the edge of the projected stimulus. Such blurring would be particularly undesirable, since it would vary with the density of the filter and the number of glass plates used.

The AC which supplied both the projection

lamps was stabilized by a Sorenson electronic voltage stabilizer. The stabilizer was found to provide sufficiently stable voltage so that luminous output was constant within ± 1% from instant to instant. Luminous output did not fluctuate with the AC due to the large heat-storage capacity of the projection lamp.

A special timing mechanism was constructed to permit stimulus presentations of accurate duration and to ensure accurate timing of the intervals between stimuli. Basically, the timer consisted of a large metal sector disc with a fixed aperture which was rotated by a synchronous motor at a rate of one rotation in 2.05 sec. At this speed, the stimulus increased to full brightness in .012 sec., remained at full brightness for .060 sec. and decreased to zero brightness in .012 sec. A stimulus presentation could be made once per rotation of the wheel, i.e., one every 2.05 sec.

It was not feasible for every possible light flash to be utilized, since it was believed that *S* would have insufficient time to observe and respond every 2.05 sec. Accordingly, a solenoid-operated shutter was mounted in the optical beam to "select" certain flashes to be presented. The solenoid opened the shutter for approximately 1 sec. The stimulus exposure was thus controlled by the sector opening in the timer wheel, not by the shutter. The solenoid was operated by a relay panel activated by a cam on the shaft of the timer so that the shutter and sector remained in synchrony. By a system of latching relays the interval between successive presentations could be varied to correspond to every fifth, sixth, or seventh revolution of the wheel, yielding intervals of 10.25, 12.30, or 14.35 sec. The interval which was used depended upon the speed of *E* in manipulating the apparatus and the comfort of *S* in responding. A microswitch on the shaft of the timer connected in series with the last relay in the panel activated an electric buzzer 1.5 sec. before the shutter was operated so that *S*s were given a regular warning signal before the stimulus was presented.

The apparatus could be used in this simple form when *S* had merely to judge whether or not the stimulus had been presented. Catch stimuli were presented by rotating the filter selector to the opaque "aperture."

In order to employ forced choice as the response indicator, it was required that the stimulus be presented in one of four temporal intervals which were close together in time and then that there be a delay before the next presentation sufficiently long so that *S* could respond. These requirements were met by adding a second wheel to the timer, which let through only one in seven flashes when an aperture in the slow wheel was aligned with the aperture in the fast wheel. When the slow wheel was in operation, the solenoid-driven shutter was disconnected. In order to vary the temporal interval in which the flash occurred, the slow wheel could be held by *E* and thus delayed 1, 2, or 3 revolutions of the fast wheel. In this way, *E* could cause a flash to be presented in one of four temporal intervals, corresponding to the first, second, third, or fourth revolutions of the fast wheel. Three revolutions of the fast wheel were then allowed for *S* to respond and for *E* to record the response and prepare for the next trial.

Four microswitches were mounted on the timer, each of which activated an electric buzzer for an instant at the beginning of one of the temporal intervals in which the stimulus could be presented. The intervals were separated by 2.05 sec., the time required for the fast wheel to rotate once. The stimulus presentation occurred for .060 sec. in the middle of the 2.05-sec. interval. Thus, the buzzer provided interval demarcation and did not overlap the stimulus presentation. A warning signal was provided 1.5 sec. before the first temporal interval by an electric buzzer which was activated by a microswitch mounted on the timer shaft.

The *E* followed a presentation schedule which governed the order of stimulus brightnesses and the temporal interval employed. In order to check on the accuracy of the manipulation of the apparatus, neon glow lamps were activated by contacts which corresponded to the positions of the two wheels of the filter selector and the position of the slow wheel with respect to the fast wheel. The correctness of the stimulus conditions was verified by *E* before each stimulus presentation was made.

Each *S* had a multi-contact selector switch with which he recorded his responses. The responses "Yes" and "No" and temporal intervals 1, 2, 3, and 4 were coded into radial positions of the selector switches. Response was made at the end of each stimulus presentation. Shields were provided so that it was not possible for *S* to have knowledge of the responses made by the others. The indicator switches excited banks of neon glow lamps near the projection apparatus so *E* could record *S*s' responses.

*Subjects.*—Four male *S*s were employed who had had no previous experience in any kind of vision experimentation. The *S*s observed together for a 2-hr. session each day, at the same time of day.

*Stimulus variables.*—Two brightness levels were studied: "zero" and .698 ft.-lambert. For "zero" brightness, the background projector was not used and the only illumination of the screen came from stray light within the room. There was not sufficient stray light for any *S* or the *E* to detect its presence. For the higher brightness, the background projector was employed at

maximum output. The brightness of the screen was measured daily with a Macbeth illuminometer and lens attachment which imaged the surface being photometered in the cube of the Macbeth. The brightnesses of the various stimuli were also measured daily, without the Wratten gelatine filters and glass plates used to reduce their brightness. The calibrations of the background and stimulus brightnesses were based upon photometric standards supplied by the Electrical Testing Laboratories of New York. Transmissions of the reduction filters were measured on an optical bench photometer, utilizing the inverse square law.

The brightness of the fixation and orientation spot was varied to suit the comfort of the *S*s. It was found that *S*s preferred the fixation spot to be about ten times its brightness-difference threshold. The *S*s were instructed to look directly at the stimuli at the high brightness and to use eccentric vision at "zero" brightness. They were instructed to select the degree of eccentric vision they each found to be most advantageous during a series of practice sessions. It was found that they chose to look eccentrically by from 10 to 20°.

*Procedure.*—The *S*s were given eight 2-hr. practice sessions in which they experienced the two brightness levels, the largest and smallest stimuli, and the two psychophysical methods. It was felt that extensive experience with the forced-choice method would have more effect upon responses made with the "Yes–No" method than vice versa. Consequently, all eight experimental conditions were completed first with the "Yes–No" method, and then all were completed with the forced-choice method. One *S* was obliged for personal reasons to be absent from several of the "Yes–No" sessions. Make-up sessions were arranged at the conclusion of the forced-choice sessions and the data were examined to determine whether there was any systematic effect which might be attributed to the intervening experience with the forced-choice method. No such effect was evident.

Eight stimulus magnitudes were employed in each experimental session. Twenty stimuli of the same magnitude were presented consecutively; then a block of 20 stimuli of a second magnitude was presented. In some "Yes–No" sessions, there were three blocks of each stimulus magnitude. In the other sessions, there were four blocks of each magnitude. The increase in the number of presentations was made possible by decreasing the length of the response periods between successive stimulus presentations. In each case, the order of presentation of the blocks of the various stimulus magnitudes was random. For the forced-choice sessions, there were always two blocks of each stimulus magnitude. Thus, the data consist of individual sessions in each of which eight stimulus magnitudes were presented a total of from 40 to 80 times. The *S*s were aware that the stimuli were blocked into groups of 20 having the same magnitude.

## Results and Discussions

The stimulus magnitudes were expressed in terms of $\Delta B$, the brightness increment added to the screen by the stimulus projector. The proportions of either "Yes" responses or of correct forced choices were determined for each stimulus magnitude in each experimental session for each *S* individually, by manual scoring of data record sheets. In the case of the forced-choice data, there is a probability of .25 of correct guesses from chance alone. Allowance for chance successes was made by the equation:

$$p = \frac{p' - c}{1 - c}, \qquad (1)$$

where $p$ = corrected proportion, $p'$ = raw proportion, and $c$ = proportion of responses which would be correct by chance alone. This formula is based upon the assumption that correct responses due to chance and correct responses due to sensory discrimination are statistically independent. Evidence supporting the validity of this assumption is reported elsewhere (2).

In the case of the proportions of "Yes" responses, if there were positive responses on catch stimuli, the same formula was employed with the proportion of positive "Yes" responses on catch stimuli utilized as an estimate of $c$. Correction of this sort was only necessary in two instances. The magnitude of these corrections was only a few per cent.

The corrected proportions were plotted against $\Delta B$ for each *S* in each experimental session and a smooth curve was fitted to the data visually. Although the curve fitting was not

TABLE 1

THRESHOLD VALUES OF $\overline{\Delta B}$ IN FOOT-LAMBERTS

| S | $\alpha$ = 4.20′ | | $\alpha$ = 10.5′ | | $\alpha$ = 28.0′ | | $\alpha$ = 77.5′ | |
|---|---|---|---|---|---|---|---|---|
| | *YN* | *T* | *YN* | *T* | *YN* | *T* | *YN* | *T* |
| | | | | $B_0$ = "zero" | | | | |
| A | $3.53\times10^{-1}$ | $3.38\times10^{-1}$ | $5.59\times10^{-1}$ | $5.77\times10^{-1}$ | $1.12\times10^{-1}$ | $.950\times10^{-1}$ | $2.74\times10^{-2}$ | $2.24\times10^{-2}$ |
| B | 2.16 | 4.66 | 3.88 | 6.45 | .720 | 1.03 | 1.73 | 2.69 |
| C | 3.52 | 4.41 | 5.72 | 7.22 | .900 | 1.09 | 2.23 | 2.49 |
| D | 3.05 | 4.31 | 5.47 | 8.22 | 1.02 | 1.10 | 1.97 | 1.97 |
| Mean | 3.07 | 4.19 | 5.17 | 6.92 | .941 | 1.04 | 2.17 | 2.35 |
| | | | | $B_0$ = .698 *ft.-lamberts* | | | | |
| A | $4.65\times10^{-1}$ | $3.84\times10^{-1}$ | $1.19\times10^{-1}$ | $.731\times10^{-1}$ | $3.82\times10^{-2}$ | $5.13\times10^{-2}$ | $1.96\times10^{-2}$ | $1.82\times10^{-2}$ |
| B | 3.10 | 3.77 | .955 | .854 | 3.70 | 3.18 | 2.01 | 1.90 |
| C | 3.95 | 3.71 | 1.03 | .974 | 3.38 | 3.18 | 1.97 | 1.79 |
| D | 4.15 | 3.78 | 1.33 | 1.23 | 4.01 | 3.24 | 1.89 | 2.15 |
| Mean | 3.96 | 3.78 | 1.13 | .947 | 3.73 | 3.68 | 1.96 | 1.92 |

restricted in advance to any mathematical function, a curve closely resembling a normal ogive was usually found satisfactory. Excluding practice runs, the present experiment represents a total of 34,000 observations by four *S*s.

Threshold values, designated $\overline{\Delta B}$, are represented in Table 1. For convenience the following symbols have been used:

$B_0$ = background brightness (ft.-lamberts),
$\alpha$ = angular subtense of stimulus diameter (minutes of arc),
*YN* = "Yes-No" data,
*T* = temporal forced-choice data.

The average values are plotted in Fig. 1. It appears that the over-all functional relations between $\overline{\Delta B}$ and $\alpha$ at the two values of $B_0$ are much the same for the two sets of data. There appear, however, to be two systematic trends. The *YN* thresholds are higher than the *T* thresholds at the higher brightness and lower at "zero" brightness. Furthermore, the difference between the *YN* and *T* thresholds at "zero" brightness appears to decrease as stimulus size is increased. The genuineness of these trends may be demonstrated most simply by expressing the data in terms of the ratio, *YN*/*T*. Values of this ratio are presented in Table 2. We note that at "zero" brightness, only 3 of the 16 ratios exceed unity. At the higher brightness, in contrast, 13 of the 16 values of the ratio exceed unity. In addition, all *S*s show a general trend for the value of the ratio to increase as stimulus size increases in the case of the "zero" brightness data.

It is possible to apply a very direct statistical argument to test the significance of both the major trends of the data. We note two characteristics of the average values of the ratio: (*a*) of the eight averages in terms of $\alpha$, the four smallest are associated with "zero" brightness, and the four largest with the higher brightness; and (*b*) the average values at "zero" brightness are regularly ordered as a function of $\alpha$. Now if there were no systematic trends present, all orderings of the data would be equally likely. We may ask, what is the probability that the data would exhibit these ordering features by chance? If the probability is small, we shall be inclined to conclude that systematic trends *are* present.

Consider (*a*) first. Denote the eight averages, arranged in order of magnitude, as $a_1$, $a_2$, . . ., $a_8$. The number of possible ordering is

$8! = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 40{,}320$. If (*a*) is to be satisfied, all orderings in which $a_1 \ldots a_4$ precede $a_5 \ldots a_8$ are admissible, and any arrangement of $a_1 \ldots a_4$ may be associated with any arrangement of $a_5 \ldots a_8$. The number of orderings of $a_1 \ldots a_4$, and likewise of $a_5 \ldots a_8$, is $4! = 24$; hence the number of combined orderings is $24 \cdot 24 = 576$.

This is not quite all, since orderings in which $a_5 \ldots a_8$ precede $a_1 \ldots a_4$ must also be considered as admissible. Since the number of such orderings is also 576, the total number of orderings satisfying (*a*) is 1152. Hence if all orderings are equally probable, the probability that (*a*) will hold is $1152/40{,}320 = 1/35$.

To analyze condition (*b*), consider the subset $a_1 \ldots a_4$. This may be ordered according to magnitude in two ways (ascending or descending) out of 24; hence the probability that $a_1 \ldots a_4$ will be ordered in magnitude is 1/12. The probability that $a_5 \ldots a_8$ will be ordered in magnitude is likewise 1/12. The probability that neither set will be ordered in magnitude is thus $(11/12)^2 = 121/144$; hence the probability that at least one set will be so ordered is $1 - 121/144 = 23/144$.

Now the probabilities of (*a*) and (*b*) are independent of each other; thus the probability of both is $1/35 \cdot 23/144 = 23/5040 = .00450$. We may conclude that the trends in the difference between *YN* and *T* thresholds are significant beyond the .5% level. Thus, we have shown that the two psychophysical procedures result in values of the threshold which are systematically different from one another as a function of each of two stimulus variables.

These results reveal that the functional dependence of brightness discrimination upon either stimulus size or background brightness may be influenced by the psychophysical procedure employed to collect the data. On the average, the observed effects are comparatively small. The functional dependence upon background brightness differs by 30% between the two methods, whereas the functional dependence upon stimulus size, at "zero" brightness, differs by 25%. Individual differences are considerably larger, however. The most extreme *S* (subject B) shows a 73% difference in terms of background brightness and a 39% difference in terms of stimulus size at "zero" brightness. These differences seem to be of sufficient magnitude to be important for current sensory theory.

There is, regrettably, no entirely direct way in which to establish which of the two methods employed gives us the "correct answer." If we studied brightness discrimination as a function of background brightness and stimulus size with a variety of procedures, we might obtain some clue as to which of the two methods provides the "cor-

TABLE 2

VALUES OF *YN*/*T* THRESHOLDS

| *S* | $\alpha = 4.20'$ | $\alpha = 10.5'$ | $\alpha = 28.0'$ | $\alpha = 77.5'$ | Mean |
|---|---|---|---|---|---|
| | | | $B_0$ = "zero" | | |
| A | 1.045 | .968 | 1.180 | 1.222 | 1.104 |
| B | .464 | .602 | .698 | .644 | .602 |
| C | .798 | .793 | .826 | .895 | .828 |
| D | .706 | .667 | .930 | 1.000 | .826 |
| Mean | .753 | .757 | .909 | .940 | .840 |
| | | | $B_0$ = .698 ft.-Lamberts | | |
| A | 1.212 | 1.631 | .745 | 1.078 | 1.166 |
| B | .822 | 1.123 | 1.163 | 1.058 | 1.042 |
| C | 1.065 | 1.058 | 1.063 | 1.100 | 1.072 |
| D | 1.098 | 1.082 | 1.238 | .878 | 1.074 |
| Mean | 1.049 | 1.223 | 1.052 | 1.029 | 1.089 |

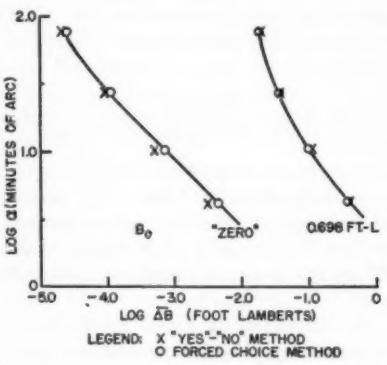


FIG. 1. Measures of two sensory functions by alternative psychophysical procedures

rect answer." Corroborative evidence from other kinds of measurements might also prove useful in this connection. The assumption that there is a "correct answer" is considered in some detail elsewhere (2).

There is one aspect of the data which provides a useful lead in this connection. The thresholds obtained at "zero" brightness were smaller for the *YN* method than for the *T* method. Data reported in an earlier paper (2) suggest that this result is not to be expected unless there are spurious "Yes" responses with the *YN* method. This possibility can be verified by carrying out an analysis of the patterns of response frequency such as was described in the earlier paper.

We begin by tallying the frequency of "No" responses as a function of order in the block of 20 stimuli of equal magnitude. The proportion of "No" responses in the first five presentations of each block is to be compared with the proportion of "No" responses in each successive five presentations. The hypothesis we are testing is that *Ss* *channelize* their responses toward the end of the block of stimuli having the same magnitude. Channelization occurs because *Ss* know that the stimuli are blocked into groups of 20 having the same magnitude. Positive channelization occurs when *Ss* "conclude" that the predominant response is "Yes." It is represented by an increase in the frequency of "Yes" responses towards the end of the block of stimuli. Negative channelization occurs when *Ss* "conclude" that the predominant response is "No." It is represented by an increase in the frequency of "No" responses towards the end of the block of stimuli. Since positive and negative channelization result in an opposite effect, it is necessary to segregate stimulus magnitudes which result in predominantly "Yes" responses from those which result in predominantly "No" responses.

TABLE 3

ANALYSIS OF FREQUENCY OF "No" RESPONSES FOR ALL *Ss*

| Stimuli | | Predominant Response: "Yes" | | | Predominant Response: "No" | | |
|---|---|---|---|---|---|---|---|
| | | *f* | *n* | *p* | *f* | *n* | *p* |
| 1st | 5 | 426 | 1555 | .274 | 725 | 935 | .775 |
| 2nd | 5 | 285 | 1555 | .183 | 730 | 935 | .781 |
| 3rd | 5 | 285 | 1555 | .183 | 734 | 935 | .784 |
| 4th | 5 | 311 | 1555 | .200 | 703 | 935 | .752 |
| Last | 15 | 881 | 4665 | .189 | 2167 | 2805 | .774 |
| | | $\chi^2 = 50.89$ $p(\chi^2) < .001$ | | | $\chi^2 = .0325$ $p(\chi^2) = .82$ | | |

The results of a tally of "No" responses for all *Ss* are presented in Table 3. All four stimulus sizes and both background brightnesses are combined. We note the drastic reduction in $p$, the proportion of "No" responses, toward the end of the series of stimuli which elicit predominantly "Yes" responses. This reduction represents positive channelization. No trend toward negative channelization is apparent.

We have tested the significance of trends in the proportions of "No" responses by the $\chi^2$ test setup as follows:

$$\chi^2 = \frac{4n\,(3\,v_1 - v_2)^2}{3(v_1 + v_2)(4n - v_1 - v_2)}, \quad (2)$$

where $v_1$ = frequency for first 5; $v_2$ = frequency for last 15; and $n$ = number of stimulus presentations in first 5 stimuli in each group of 20. When this setup is employed, there is 1 *df*.

As indicated in Table 3, the trend representing positive channelization is significant far beyond the .1% level. There is no significant trend among

the data when the predominant response is "No."

The presence of significant positive channelization confirms the result of the earlier experiment (**2**). It is not unreasonable that negative channelization does not appear, as it did not in the earlier experiment. The *S*s knew that only one group of catch stimuli occurred for each eight groups of stimuli. Hence, they could not tolerate consistent "No" responses for very many groups of stimuli.

These results emphasize that the "Yes-No" method utilizing grouped stimuli offers *S* the opportunity for invalid modes of response. We must remark in passing that the "Yes-No" method, here again shown to be invalid, is very nearly the method required by the neural quantum theorists (**3, 4**) in order for their theory to be applicable. The only difference is that we have included catch stimuli, whereas the proponents of the neural quantum theory have not. One would suppose that positive channelization would be even more apparent in the absence of catch stimuli.

Perhaps the most interesting aspect of this matter is the possibility that response channelization may distort psychophysical data into the linear form predicted by the neural quantum theory. It seems reasonable to expect that channelization will be more pronounced the farther the stimulus magnitude is removed in either direction from the threshold. Positive channelization would spuriously increase probabilities near unity. Negative channelization would spuriously decrease probabilities near zero. The general effect would be to distort data which were actually represented by a normal ogive into the linear form predicted by the neural quantum theory. The data obtained in the present experiment have not been considered sufficiently numerous to justify an analysis of the extent of channelization as a function of stimulus difference from threshold. Of course, as noted, both positive and negative channelization are required to distort the data at both ends. Although there was no evidence of negative channelization in our data, it seems possible that when catch stimuli are not used, as was the case in the experiments purported to support the neural quantum theory, the likelihood of negative channelization is increased, since *S*s no longer feel that only a few sets of stimuli should elicit consistent "No" responses.

As noted elsewhere (**2**) the neural quantum theory may be investigated best in terms of the mean-slope relations obtained. The present data cannot be usefully analyzed in these terms because only eight stimulus magnitudes were utilized to define the psychophysical curve.

Since significant positive channelization has been shown to occur with the "Yes-No" method, we may well investigate how well the extent of this invalid mode of response correlates with the extent and direction of difference between $YN$ and $T$ thresholds. As a crude measure of the extent of channelization, we define the channelization index,

$$I = v_1/v_2, \qquad (3)$$

where $v_1$ and $v_2$ have the same meanings as in formula (2). Values of $I$ greater than unity represent positive channelization, whereas values less than unity represent negative channelization. Average values of $I$ for all four *S*s are presented in Table 4 for each of the four stimulus sizes and the two background brightnesses. We note that there is evidence of positive channelization for each of the eight stimulus conditions. There is evidence of negative channelization for

TABLE 4

VALUES OF THE CHANNELIZATION INDEX FOR THE PREDOMINANT RESPONSE IN ALL *Ss*

| $\alpha$ | $B_0$ = 0 ft.-lamberts | | $B_0$ = .698 ft.-lamberts | |
|---|---|---|---|---|
| | "Yes" | "No" | "Yes" | "No" |
| 4.2′ | 1.52 | .99 | 1.74 | .83 |
| 10.5′ | 1.49 | 1.09 | 1.04 | 1.00 |
| 28.0′ | 1.38 | 1.03 | 1.37 | 1.02 |
| 77.5′ | 1.58 | 1.07 | 1.45 | .99 |
| Mean | 1.49 | 1.04 | 1.38 | .94 |

only three of the eight stimulus conditions.

Since only positive channelization is significant, we investigated the notion that the extent of channelization influences the extent of difference between the *YN* and *T* thresholds by correlating values of *I* for predominantly "Yes" responses with values of the *YN*/*T* ratio contained in Table 2. The value of $\rho$, the rank-order correlation coefficient, is $-.45$. Although not clearly significant, this correlation is suggestive that positive channelization is part of the cause for the differences between the *YN* and *T* threshold. The correlation would be expected to be negative. We would expect that positive channelization would spuriously reduce the *YN* threshold in comparison with the *T* threshold, thus decreasing the size of the *YN*/*T* ratio as the value of *I* increased. Other factors than positive channelization must be involved and we can only guess as to their identity.

It is to be emphasized that the opportunity for response channelization is not present with the forced-choice method. In this case, the stimulus magnitudes are ordered into blocks of 20 having the same value, but the correct answers, the temporal interval occupied by the stimulus, are randomized. Thus, there is no direct way in which knowledge of the ordering of stimulus magnitudes may lead to invalid modes of responses.

## SUMMARY AND CONCLUSIONS

It has been shown that measurements of thresholds of brightness discrimination depend upon psychophysical procedure in a manner which varies as a function of stimulus size and background brightness. The dependence upon the stimulus variables is significant and of a magnitude at least in some cases sufficiently large to be important for contemporary sensory theory. The sole difference in the two procedures studied was the response employed to indicate the presence of discrimination. In one case, phenomenal report was employed; in the other, forced choice was employed. The stimuli were ordered into groups of 20 in succession having the same magnitude.

It was shown that *Ss* exhibited an invalid mode of response, termed *positive channelization*, when phenomenal report was employed. This response tendency consists of spuriously increasing the frequency of "Yes" responses toward the end of a group of 20 stimuli of the same magnitude, providing the stimuli elicit more than 50% "Yes" responses. The extent of positive channelization was shown to correlate $-.45$ with the differences between the thresholds obtained by the two psychophysical procedures as a function of stimulus size and background brightness.

It is concluded that the procedure employed to measure thresholds can distort functional relations studied by psychophysical methods. It is concluded further that the use of phenomenal report with grouped stimuli can lead to invalid data. This combination of psychophysical procedures has been specifically advocated by the

proponents of the neural quantum theory. The question arises to what extent invalid modes of response arising from the use of this psychophysical procedure have led to spurious evidence for the neural quantum theory.

(Received March 16, 1952)

## References

1. Blackwell, H. R. Contrast thresholds of the human eye. *J. opt. Soc. Amer.*, 1946, **36**, 624–643.

2. Blackwell, H. R. *Psychophysical thresholds: Experimental studies of methods of measurement.* Engr. Res. Inst. Bull., Univer. of Michigan, 1952.

3. Miller, G. A., & Garner, W. R. The effect of random presentation on the psychometric function: Implications for a quantal theory of discrimination. *Amer. J. Psychol.*, 1944, **57**, 451–467.

4. Stevens, S. S., Morgan, C. T., & Volkmann, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.*, 1941, **54**, 315–335.

# A CONSTRUCTION AND ATTEMPTED VALIDATION OF SENSORY SWEETNESS SCALES[1]

SHELTON MacLEOD

*Hobart College*

The construction of sensory scales based on supraliminal estimates of sensory magnitude has been the concern of a number of investigators in recent years. Following the establishment of two auditory scales for pitch and loudness by Stevens (**10, 11**), measurement of this type has been applied to at least three other sense modalities. Harper and Stevens (**6**) have constructed a proprioceptive scale of subjective heaviness, Lewis (**7**) has established four taste scales, and Hanes (**5**) has developed a visual scale of brilliance.

Much of the impetus behind these and other attempts to scale sensation in terms of direct numerical estimates has undoubtedly arisen from recent evaluations of the relative merits of various scaling techniques. Stevens (**12**), in particular, argues that scales derived in this manner can be classed as interval scales, being endowed with the same properties of additivity that are enjoyed by the balance and meterstick. Such a possibility is intriguing and, if indeed true, could lead to theoretical and practical consequences of considerable importance. Studies by Beebe-Center and Waddell (**1**), and Hanes (**5**) would appear to have lent some support to Stevens' contention.

The present paper is an attempt to assess the validity of sweetness scales based on fractional judgments of one-half. The method of validation here proposed may be briefly outlined as follows:

1. Two fractionation functions are to be obtained for two physically different sweet agents; namely, sucrose and glucose.

2. A function relating equally sweet concentrations of the two sugars is also to be determined.

3. If sweetness scales for the two sugars based on the fractionation functions are valid interval scales, then it has to be demonstrated that, after such scales are anchored together along a common dimension of sweetness, any pair of sucrose and glucose concentrations eliciting the same number of sweetness units must always prove to be equally sweet. In other words, if the scales are valid, then one should be able to predict the obtained sweetness matches from them.

## METHOD

### *Method Involving Sucrose and Glucose Fractionation*

*Apparatus and situation.*—All *S*s were seated at a table by a sink and were provided with an ample supply of tap-water rinse.[2] A black curtain was suspended across the center of the table, the curtain having a flatlike opening

[1] The research reported in the present experiment was carried out at Brown University in partial fulfillment of the requirements of the degree of Doctor of Philosophy. It received partial support from Project Nr 140–721, Contract N7-onr-35804 between Brown University and the Office of Naval Research. Acknowledgment should be given in connection with it to Dr. Carl Pfaffmann who directed the research and furnished valuable advice and assistance in its execution. The writer also wishes to express his indebtedness to Mr. Ogden Lindsley, Mr. Anthony Davids, and Mr. Preston Abbott who served as experimental *S*s.

[2] In the sucrose fractionation procedure for subject O. L. distilled, rather than tap water, was used. A check experiment gave assurance that this minor lack of standardization was insignificant.

just large enough to permit an exchange of 1-oz. medicine glasses (which always contained the experimental solutions) between *S* and *E*, but not large enough to reveal any of the experimental manipulations to *S*. All of *S*'s judgments were made after tasting the comparison solutions, and were referred to an interval of sweetness bounded by the respective sensations arising from the standard and the rinse water. The *S*s judged, in terms of such an interval, whether the sensation elicited by the comparison solution was greater or less than halfway along this interval. No responses of "equal" were permitted.

The *E* dispensed single 15-cc. amounts of either standard or comparison solution into the 1-oz. medicine glasses at his end of the table. Upon receiving the comparison solution *S* held it in his mouth until it reached a peak intensity, then formed his judgment and expectorated the solution into the sink. Immediately thereafter he rinsed his mouth thoroughly, wrote down his judgment in the form of plus or minus, and awaited the next solution.

The standard solution was given by *E* at the start of any session and was thereafter announced whenever presented. All presentations were separated by an interval of 1 min.

*Specification of sugar solutions.*—One commercial brand of cane sugar (Revere) was employed for all sucrose solutions. This was tested as 99% pure in a polarimeter.[3] All glucose used was of a USP high grade of purity. Solutions were always freshly prepared with a tap-water solvent made up to a required volume percentage concentration; i.e., number of grams of solute in 100 cc. of solution. The solutions and rinse water were allowed to equilibrate to room temperature before the start of any experimental session. This temperature averaged about 74° F. and had an extreme range of 68° to 80° F.

*Subjects.*—The *S*s were three male psychology students at Brown University. All were moderate smokers.

*Procedure.*—Before an adequate design of the fractionation experiment could be made, two preliminary checks were first necessary to determine and eliminate from consideration ranges of sugar concentration which might introduce either a disturbing element of viscosity or an undesirable subthreshold effect. The absolute threshold was therefore roughly evaluated for the three *S*s, as was the threshold of viscosity. As a result of these checks, a working range of 1 to 32% sucrose and 2 to 32% glucose appeared advisable.

The psychophysical procedure decided upon was a modified method of constant stimulus differences in which one standard and four equally-spaced comparison solutions were given equally often, each following each two times. The second half of this rotation scheme reversed the sequence of the first half. One initial and one reversed sequence constituted a daily 40-min. session. Thirty-two judgments were made during this time, eight judgments per comparison stimulus.

Six sucrose and five glucose standards were used. The standards for each sugar were separated by ¼-log steps of per cent concentration, being: 1.8, 3.2, 5.6, 10.0, 18.0, 32.0 for sucrose; and 3.2, 5.6, 10.0, 18.0, 32.0 for glucose.

The order in which the standards were introduced in 22 successive determinations was set by a counterbalanced design, which both alternated the two sugars[4] and staggered the intensity levels of successive standards. This order was as follows for the first half of the determinations: 10% sucrose, 3.2% glucose, 32% sucrose, 18% glucose, 1.8% sucrose, 10% glucose, 3.2% sucrose, 32% glucose, 18% sucrose, 5.6% glucose, and 5.6% sucrose. The order for the second half of the determinations was the exact reverse of the above order.

At least three sessions were required to determine a halfway judgment for each standard:

1. A brief 20-min. preliminary session during which widely spaced comparison solutions were used to establish a broad frame of reference and bracket the range of uncertainty.

2. An initial 40-min. experimental session immediately following the practice session.

3. A repetition of the experimental session which followed according to the counterbalanced schedule. In several instances it was necessary to follow up one or both experimental sessions involving a given determination with an additional session or sessions to clarify the results.

*Computational procedure.*—The procedure by which the most probable halfway concentrations were arrived at requires some clarification.

The first, and by far the more usual, procedure was similar to the normal graphic solution for computing absolute limens as reported by Guilford (**4**). Following this method sugar concentrations corresponding to halfway judgments were computed graphically as medians on probability paper from a line fitted (by the method of averages) to the percentages of greater-than-half judgments. A majority of such fits were good when sufficiently large numbers of judgments were involved.

In about 5% of the cases, however, proba-

[3] Acknowledgment is hereby made to Dr. L. B. Clapp of the Brown University Department of Chemistry for his assistance in such a test.

[4] In the case of subject O. L. the two sugars were fractionated separately, but it seems unlikely that this contributed to an experimental error.

bility plots could not be made, since percentage judgment values of 100 or zero had to be included. In such cases an otherwise identical method of computation was carried out on arithmetic grid paper. A series of checks showed this to be a defensible alternative to the more orthodox procedure.

If it had been possible to employ the same comparison series for all the fractionation judgments of a given standard, then one halfway concentration for that standard was arrived at from pooled data.

If, on the other hand, the use of different comparison series became necessary, a separate computation was made for each different series, and the most probable halfway judgment was taken as a weighted mean of the values from these separate determinations. Weighting was in terms of the total number of judgments involved. For example, suppose two halfway judgment values, based respectively on 144 and 96 judgments, were to be averaged by such weighting; the first value would then be multiplied by 3, the second by 2, and this total divided by 5.

It appeared necessary to discard results from about 5% of the daily determinations. However, the following liberal rejection criteria were adhered to: (*a*) Inversion had to be present to the degree that the inverted comparison stimulus was displaced by more than two greater-than-half judgments from both of its adjacent comparison stimuli; (*b*) Failure to reach the 50% line on probability paper at either the highest or lowest comparison stimuli had to be shown. Otherwise, any series of judgments was accepted, even when its use precluded probability plots. The liberal criteria of rejection were necessary in view of the fact that relatively few judgments could be made in a daily session; hence, failure to meet strict criteria was likely.

### *Method Involving Sucrose and Glucose Equal-Sweetness Matches*

After completion of the fractionation series, the same *S*s were required to equate sweetnesses of sucrose and glucose at four widely spaced intensity levels. Throughout these sessions each of four comparison solutions of one sugar was judged to be either sweeter-than or less-sweet-than a standard solution of the other sugar. Otherwise the psychophysical procedure was identical to that pursued in the fractionation series.

During the first half of this experiment standard solutions of 2, 5, 10, and 25% sucrose were prepared and each was used in conjunction with four comparison glucose solutions. The latter were suitably chosen (again on the basis of preliminary investigation) to permit computation of a point of subjective equality. The sucrose standards were introduced on successive days in the following staggered order of strength: 10%, 2%, 25%, and 5%.[5]

Four corresponding glucose matches were calculated for each *S* from these standards. The computational steps followed in arriving at these equally sweet glucose concentrations were identical to those employed in assigning most probable halfway values.

In the second half of the experiment the four matching glucose solutions were introduced in counterbalanced order as standards, and were then judged in relation to appropriately selected sucrose comparison stimuli. Sucrose matching values were again computed for each *S* and the final corrected sucrose match at each intensity level was taken as the mean of this matching value and the value of the original sucrose standard.

## Results

### *Qualitative Results*

Ideally, it should be assumed that the two sugars used elicited the identical sensation. Such was not quite the case. The *S*s reported that glucose had a fruity component which increased with concentration. Concentrated glucose solutions were also reported to be less pleasant in taste than concentrated sucrose. In any case, it could not be presumed that we were dealing with indiscriminable qualitative differences between the two sweet agents.

This fact, however, was not too surprising. To find two sweet substances that were clearly indistinguishable would be unusual in view of reports in the literature (2). Furthermore, the fact that slight qualitative differences did occur would probably only have the effect of slightly increasing constant errors and variability which were already manifest from other sources.

[5] Subject P. A. was not available for the 5% sucrose determination; hence, his most probable glucose match for that value was estimated by interpolation.

TABLE 1
SUMMARY TABLE OF SUCROSE AND GLUCOSE FRACTIONATION

| % Concentration of Standard | Sucrose Fractionation | | | | | | | Glucose Fractionation | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | O. L. | | A. D. | | P. A. | | Mean | O. L. | | A. D. | | P. A. | | Mean |
| | Series | $I_H$* | Series | $I_H$ | Series | $I_H$ | $I_H$ | Series | $I_H$ | Series | $I_H$ | Series | $I_H$ | $I_H$ |
| 1.8 | 2 | 1.08 | 2 | 1.18 | 2 | 1.10 | 1.12 | | | | | | | |
| 3.2 | 3 | 1.72 | 2 | 1.65 | 2 | 1.70 | 1.69 | 2 | 2.46 | 2 | 2.40 | 2 | 2.42 | 2.43 |
| 5.6 | 4 | 2.68 | 2 | 2.05 | 2 | 3.15 | 2.63 | 2 | 3.79 | 3 | 3.52 | 2 | 3.44 | 3.58 |
| 10.0 | 5 | 4.46 | 3 | 3.80 | 3 | 4.19 | 4.15 | 2 | 5.36 | 4 | 4.03 | 2 | 4.16 | 4.51 |
| 18.0 | 8 | 5.67 | 2 | 4.22 | 3 | 7.46 | 5.78 | 2 | 8.59 | 2 | 5.99 | 2 | 8.50 | 7.69 |
| 32.0 | 4 | 7.49 | 2 | 5.49 | 2 | 7.78 | 7.07 | 2 | 11.55 | 3 | 9.28 | 2 | 11.72 | 10.85 |

* $I_H$ refers to per cent concentration judged half.

*Quantitative Results*

*Fractionation.*—A summary of the results from the fractionation experiments is given in Table 1. Both tables reveal consistent individual differences in the $I_H$ columns (denoting per cent concentrations judged half). Subject A. D. had the general tendency to halve his sweetness intervals at lower concentrations, while O. L. and P. A. fractionated roughly at the same somewhat higher levels.

Figure 1 shows average $I_H$ and $I$ values for both sucrose and glucose plotted in logarithmic units. Straight lines fitted to these points by the method of averages appear to be reasonable descriptions of their trend. The equation for the sucrose line is: $log\ I_H = .681\ log\ I - .110$; while the glucose equation is: $log\ I_H = .651\ log\ I + .043$.

*Equal sweetness.*—Table 2 shows the most probable average percentage values of sucrose and glucose judged equally sweet. Dahlberg and Penczek (3), in particular, have reported matches which closely parallel these. Other workers (2, 8) have not obtained glucose values as low as these in matches at high levels of intensity. Our results, nonetheless, are consistent with the general finding of a progressive decrease in the glucose concentration required for matches.

When plotted logarithmically, the values of Table 2 are well fitted by a straight line. This line, shown in Fig. 2, may be described by the equation: $log\ I_G = .681\ log\ I_S + .439$ where $I_G$ and $I_S$ are the respective per cent concentrations of glucose and sucrose required for a sweetness match.

## DISCUSSION

Having obtained the required three sets of sweetness relationships, we are now ready to perform the proposed test of validity. The logical development of the present method of making such a test will be explained step by

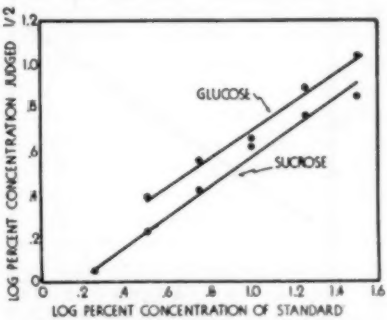


FIG. 1. Fractionation functions for sucrose and glucose (mean of three *S*s)

TABLE 2

PER CENT CONCENTRATIONS OF SUCROSE AND GLUCOSE JUDGED TO BE EQUALLY SWEET (MEAN OF THREE *Ss*)

| Sucrose Concentration | Equally Sweet Glucose Concentrations |
|---|---|
| 1.9 | 4.3 |
| 4.9 | 8.1 |
| 9.5 | 12.3 |
| 23.0 | 24.0 |

step in the following paragraphs. It should be emphasized that this development is experimentally grounded in the findings reported above.

Before proceeding with such a discussion, let us assemble the empirically determined equations which will be used to clarify the sequence of steps to follow:

$$\log I_{Glucose} = .681 \log I_{Sucrose} + .439 \quad (1)$$

For sucrose,

$$\log I_H = .681 \log I - .110 \quad (2)$$

For glucose,

$$\log I_H = .651 \log I + .043 \quad (3)$$

As a first step, let us make the not unreasonable assumption that empirical Equation (1) (the matching equation) is true and never question its validity. Then, for the moment, let us assume that Equation (2) (the sucrose fractionation equation) is also true. Such a pair of assumptions fixes the constants of a hypothetical glucose equation which, if used in conjunction with Equation (1), will always predict equal-sweetness matches. To clarify this statement let us consider the following illustration.

FIG. 2. An equal-sweetness function for sucrose and glucose (mean of three *Ss*)

Equation (2) for sucrose tells us that when *log I* equals .92, then *log* $I_H$ equals .51. Similarly, when *log I* equals .51, then *log* $I_H$ equals .23. Equation (1) tells us that .92, .51, and .23 log-concentration sucrose are, respectively, equal to 1.06, .79, and .56 log-concentration glucose. It can be seen that, if Equations (1) and (2) are valid, two pairs of glucose *log I–log* $I_H$ coordinates become fixed; namely, 1.06,. 79; and .79, .56. A hypothetically valid glucose fractionation function may be drawn through these coordinates. We will refer to the equation describing this function as Equation (4), but for the present leave it unspecified.

Let us suppose we have experimentally obtained Equation (2) and Equation (4), but have knowledge of only one sweetness match showing .92 log concentration of sucrose equal to 1.06 log concentration of glucose. The log values of at least two other sweetness matches: .51 sucrose and .79 glucose, in addition to .23 sucrose and .56 glucose would, in such case, be predicted solely in terms of the two given $I_H$–$I$ functions. A linear equal-sweetness function drawn through these points would have been identical to Equation (1).

Since fractionation lines such as the ones given by Equations (2) and (4) are able to predict the equality line given by Equation (2), we will refer to them as prediction lines.

The Equation (4) discussed above

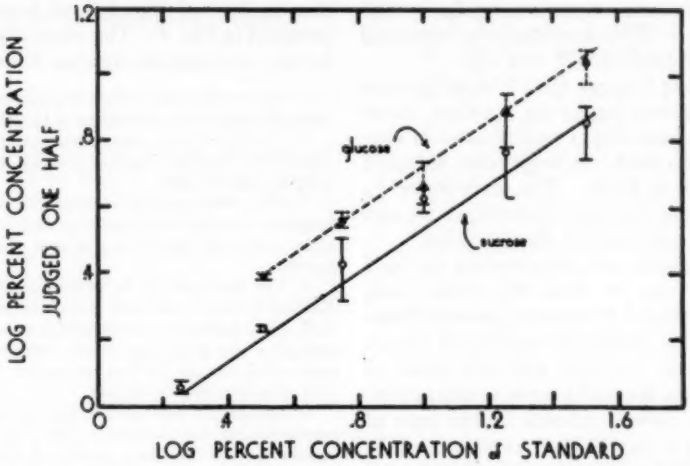


FIG. 3. Hypothetically valid fractionation functions for sucrose and glucose. In almost every case they cross bands of intersubject variability plotted above and below the same fractionation values represented in Fig. 1.

(hypothetical glucose formula) was found in the present experiment to be: $log\ I_H = .681\ I + .065$. It formed prediction lines with Equation (2).

Since it was equally logical to assume that Equation (3) (the obtained glucose-fractionation equation) was valid, a hypothetical sucrose equation: $log\ I_H = .651\ log\ I - .162$ (Equation 5) was similarly derived. Equations (3) and (5) also formed prediction lines.

Two observations were made concerning the above pairs of prediction lines: (*a*) Since there were two pairs, the obtained fractionation functions could not serve to predict the equal-sweetness matches without error; i.e., the obtained formulae and the hypothetical formulae necessary for perfect prediction (Equations 2–5 for sucrose and 3–4 for glucose) differed in the case of each sugar; (*b*) Since each of the two pairs was parallel (had identical slope constants) this property (parallelism) also seemed necessary for predictive purposes.

On the basis of these observations, one pair of parallel adjusted prediction lines was constructed for the two sugars. These lines bisected the angle formed by the slopes of the lines described by Equations 2–5, as well as the angle formed by the slopes of the lines described by Equations 3–4. Hence their construction was consistent with the assumption that both obtained fractionation equations were equally in error. Such "best-fitting" prediction lines were to be used in assessing the validity of the fractionation results in the present experiment. These lines were constructed in the following manner: (*a*) The arithmetic mean was taken for the slope constants in Equations (3), the obtained glucose equation, and (4), the hypothetical glucose equation. (*b*) The geometric mean was taken for the intercepts of the same two equations. (*c*) The adjusted glucose equation, $log\ I_H = .666\ log\ I + .054$, was obtained by using the above mean slope and intercept in its formula. (*d*) The ad-

justed sucrose equation, $log\ I_H = .666\ log\ I_G - .136$, was similarly obtained from Equations (2) and (5).

Figure 3 shows the obtained glucose and sucrose $log\ I_H$–$log\ I$ values transferred from Fig. 1 and plotted in conjunction with the respective adjusted prediction lines. The inter-$S$ $log\ I_H$ range has also been inserted above and below each one of these values.

Insofar as our obtained $log\ I_H$–$log\ I$ values can be fitted by these lines, such values will have established themselves as useful in predicting equal-sweetness matches and can serve as the basis for valid scale construction.

The correspondence in the case of glucose is particularly promising since four of the five obtained points fall approximately on the prediction line. One $log\ I_H$ value, for $log\ I$ of 1.0, is noticeably displaced, but the band of variability around it crosses the prediction line.

In the case of sucrose, $log\ I_H$ percentages for $log\ I$ values of .75, 1.00, and 1.25 deviate somewhat from the adjusted prediction line; but, only in the case of the $log\ I_H$ corresponding to $log\ I$ of 1.00, does the inter-$S$ $log\ I_H$ range fail to cross the line.

Scales based on the adjusted prediction lines of Fig. 3 have been constructed in Fig. 4. The steps involved in the construction were as follows:

1. It was first assumed that a sucrose $log\ I$ value of 1.40 elicited a sweetness of 14 scale units.
2. This $log\ I$ value was substituted into the adjusted sucrose fractionation equation, yielding a $log\ I_H$ value of .80.
3. This value was, by the nature of the half-judgment determining it, taken as eliciting seven scale units and resubstituted into the same equation.
4. The resultant $log\ I_H$ value was taken as eliciting 3.5 scale units and again resubstituted.
5. Such a process was repeated once again to provide a descending geometric series of four scale steps elicited by four respective log per cent concentrations of sucrose.
6. One log per cent glucose and log per cent sucrose match was assumed on the basis of the equal-sweetness function; namely, 1.39 log % glucose equals 1.40 log % sucrose.
7. This glucose concentration in terms of the original sucrose definition also elicited a sweetness of 14 scale units.
8. The adjusted glucose fractionation equation was then employed to scale glucose by the same process as was indicated for sucrose.
9. Anti-logarithms of the physical values of sucrose and glucose found in the above steps were plotted against the respective scale units of sucrose and glucose which they elicited.

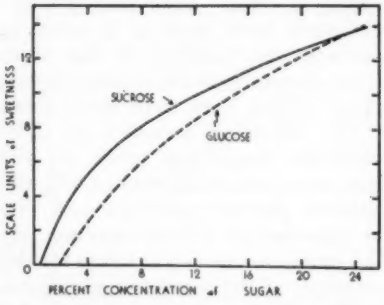


FIG. 4. Scale units of sweetness for glucose and sucrose plotted against per cent concentration of sugar. The scales have been constructed from adjusted prediction lines.

The functions relating scale units to physical concentrations are both negatively accelerated as can be seen in Fig. 4. It is of interest to note that the relationship between the sensory and physical dimensions in the case of glucose is not far from linear up to a concentration of 8%.

Also of interest is the fact that these scales always meet with the criterion of validity proposed in the introduction since concentrations of the two sugars eliciting the same number of sweetness units will always fall along the obtained equal-sweetness line; and, since the hypothetical fractionation lines used in constructing these scales were not far removed from the obtained fractionation values, it seems reasonable to assume in this instance that the method of fractionation has

served as a tolerably sound basis for valid interval scale construction.

## Conclusions

The following conclusions may be drawn from the present experiment:

1. Log concentrations of sucrose and glucose judged half bear a linear relationship to log concentrations of sucrose and glucose standards. In the case of glucose the equation describing this relationship is $log\ I_H = .651\ log\ I + .043$; in the case of sucrose the equation is $log\ I_H = .681\ log\ I - .110$.

2. Log concentrations of glucose bear a linear relationship to equally sweet log concentrations of sucrose as given by the equation: $log\ I_{Glucose} = .681\ log\ I_{Sucrose} + .439$.

3. Inasmuch as hypothetically valid fractionation functions for both sugars could be fitted (within *S*-to-*S* variability) to the obtained fractionation values, it would appear reasonable to assume that the method of fractionation did, in this instance, provide data for valid interval scale construction.

(Received April 3, 1952)

## References

1. Beebe-Center, J. G., & Waddell, D. A general psychological scale of taste. *J. Psychol.*, 1948, **26**, 517–524.
2. Cameron, A. T. The taste sense and the relative sweetness of sugars and other sweet substances. *Scientific Reports of the Sugar Research Foundation*, New York, 1947, No. 9.
3. Dahlberg, A. C., & Penczek, E. S. The relative sweetness of sugars as affected by concentration. *N. Y. St. Agric. Exp. Sta. tech. Bull.*, 1941, No. 258.
4. Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
5. Hanes, R. M. The construction of subjective brightness scales from fractionation data: A validation. *J. exp. Psychol.*, 1949, **39**, 719–728.
6. Harper, R. S., & Stevens, S. S. A psychological scale of weight and a formula for its derivation. *Amer. J. Psychol.*, 1948, **61**, 343–351.
7. Lewis, D. R. Psychological scales of taste. *J. Psychol.*, 1948, **26**, 437–446.
8. Lichtenstein, P. E. The relative sweetness of sugars: sucrose and dextrose. *J. exp. Psychol.*, 1948, **38**, 578–586.
9. Reese, T. W. The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. *Psychol. Monogr.*, 1943, **55**, No. 3 (Whole No. 251).
10. Stevens, S. S. A scale for the measurement of a psychological magnitude: Loudness. *Psychol. Rev.*, 1936, **43**, 405–416.
11. Stevens, S. S., Volkmann, J., & Newman, E. B. A scale for the measurement of the psychological magnitude pitch. *J. acoust. Soc. Amer.*, 1937, **8**, 185–190.
12. Stevens, S. S. On the theory of scales of measurement. *Science*, 1946, **103**, 677–680.

# "ABNORMAL FIXATION" AND LEARNING[1]

HARDY C. WILCOXON

*The Psychological Corporation*

Recently Maier has brought together his evidence for "abnormal fixation" (**3**) and reaffirmed his stand that the behavior cannot be accounted for by any current learning theory (**4**). Maier attributes fixation to frustration which is said to arise either from severe punishment, confusion over equally undesirable alternatives in an insoluble problem, or a combination of the two. Once an organism undergoes frustration, his behavior is said to take on an abnormally rigid character qualitatively different from learned habits, and not subject to change by conventional learning techniques. The present research is an attempt, from an S–R learning theory point of view, to throw new light on this complex issue.

The training condition which Maier has called the Insoluble Problem, and which provides the basis for most of his major conclusions, combines many complex factors known to be important determiners of behavior. In this situation rats are forced with an air blast to respond on the Lashley jumping stand, even though the stimulus cards, which are switched from left to right in the windows of the apparatus, are locked in a chance order. The chance order of the cards guarantees that the animal will, on the average, receive punishment on half of his trials (hitting a locked card and falling into a net) and reward on the other half (knocking over an unlocked card and receiving food on the rear platform), regardless of whether he remains variable in his choices or chooses a particular card or window consistently. The air blast insures that some response will be made quickly. Thus, Maier's Insoluble Problem combines the following important factors: (*a*) Punishment—both at the locked cards and in the form of an air blast used to force jumps; (*b*) Lack of differential reinforcement—which makes a nonselective learning condition; (*c*) A complex partial-reinforcement factor—which arises from the discontinuous nature of the reinforcing conditions, i.e., alternating reward and punishment of the same response.[2]

Maier believes that factors (*a*) and (*b*) are the important ones leading to frustration and the consequent fixation of behavior. Sheer punishment[3] is said to frustrate if received in large amounts, while lack of differential reinforcement is said to confuse the *S* and thereby create additional stress (**3**, p. 69). He does not deal explicitly with factor (*c*), a learning variable

[1] Based on a dissertation presented to the faculty of the Graduate School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to thank Dr. Fred D. Sheffield, under whose direction the research was conducted, and Doctors Irvin L. Child, Neal E. Miller, and Carl I. Hovland, who served on the Advisory Committee. A preliminary version was read at the Sept., 1951 meeting of the American Psychological Association.

[2] The generalization decrement theory of partial-reinforcement effects (**8**) is applicable to this situation as well as the more typical one where one type of reinforcement is involved.

[3] Maier lightly dismisses the reinforcing property of escape from punishment which has been an important consideration in S–R treatments of learning for 20 years (**10**). Although it was not the purpose of the present research to investigate air-blast punishment, the writer recognizes its important role in the persistence of the behavior described by Maier.

with very predictable effects which have been well documented (**2**).

In order to determine the direction an S–R explanation of "abnormal fixation" should take, it is necessary to determine the separate as well as the combined effects of some of the factors in the Insoluble Problem. If it turns out that lack of differential reinforcement is a critical factor which confuses the rat and cuts short his normal period of variability in a problem (**3**, p. 69), an S–R explanation will have to proceed along relatively unexplored lines. Other things being equal, S–R theory predicts greater variability with nondifferential reinforcement. However, if this factor does not contribute to fixation, explanation can proceed along the well-established lines of S–R versions of partial-reinforcement theory (*cf.* **2**), and Maier's less parsimonious theory will appear unnecessary.

Two experiments are reported. In the first, the factor of nondifferential reinforcement is isolated from the factor of partial reinforcement. In the second, two learning procedures are demonstrated which allow *S*s to unlearn fixations.

## Experiment I

Table 1 summarizes the logic of the experiment. The group comparisons which can be made in terms of ability to learn the subsequent test problem are as follows:

1. Group I versus II compares groups which are alike in that their initial training was under selective-learning conditions. They differ in that one group was trained under partial-reinforcement conditions, the other under continuous. The comparison will reveal the effect of partial reinforcement.

2. Group II versus III compares groups which are alike in having had partial-reinforcement conditions in initial training, but which differ in that one of the initial-training conditions was selective while the other was nonselective. The comparison will therefore show the effect of nondifferential reinforcement.

3. Group I versus III is a comparison between one group whose initial training was under selective, continuous-reinforcement conditions and another whose first training was under nonselective, partial–reinforcement conditions. Consequently, the comparison will reveal the combined effects of nondifferential reinforcement and partial reinforcement. This comparison is Maier's basic design.

### Method

*Apparatus.*—The jumping stand was patterned after the one pictured and described in Maier's book (**3**). The stimulus cards were a white card with a solid black circle (hereafter called BC) and a black card with a solid white circle (WC). An air blast of between 12 and 15 lb. was connected to a ¼-in. nozzle at the rear of the jumping platform. An electric timer was used to record latencies to the nearest second.

*Subjects.*—The *S*s were 45 male albino rats between 3 and 4 months old and experimentally naive at the outset of training.[4] They were kept on a 24-hour hunger cycle and fed 10 gm. of ground laboratory chow (dry weight) mixed with water.

*Preliminary training.*—Preliminary training in jumping to the stimulus cards closely followed that described in detail by Maier, Glaser, and Klee (**5**).

*Group assignments.*—Rats were assigned randomly to the experimental groups after they had completed preliminary training, insuring that no

[4] A number of rats which were started in preliminary training died of a kidney infection during the second week of the experiment. When the illness was discovered, all rats in the colony were given two intraperitoneal injections of penicillin per day for three days. The dosage was 50,000 units of crystalline penicillin in aqueous solution. Rats were not run on these days but were fed at their regular times. The rats which recovered and were continued in the experiment were in good health.

TABLE 1

DESIGN OF EXPERIMENT I

| | Groups | Learning Condition | Distribution of Reinforcement |
|---|---|---|---|
| Training procedure during experimental training | I Continuous reinforcement | Selective | Continuous reward for correct response (one side always unlocked, the other always locked) |
| | II Partial reinforcement | Selective | Partial reinforcement of correct response (correct side unlocked on 50% of trials, other always locked) |
| | III Insoluble problem | Nonselective | Partial reinforcement for either position or either card (both sides and cards locked and unlocked 50% of trials) |
| Test procedure after experimental training | I<br>II<br>III | All groups received a problem in which the response which always led to food required a shift from the initial response. Rats with position habits received a visual discrimination problem. Rats with card habits received a problem in which a position habit invariably led to food. | |

systematic differential treatment of the three groups could occur in preliminary training.

*Experimental training.*—The different conditions for the three groups are described in Table 1. However, certain training factors were common to all experimental groups.

The air-blast schedule used by Maier was adhered to throughout experimental training and test training. Three intensities were used depending upon how hesitant the rat was. If *S* jumped within 30 sec. after being placed on the stand, no air was applied; from 31 to 60 sec., mild air was applied; from 61 to 90 sec., medium air; and from 91 to 300 sec., full air blast. The intensities "mild" and "medium" were subjectively judged by *E* to be evenly spaced between no air and full air. Places on the valve corresponding to these intensities were marked and used consistently.

All groups received 200 trials in which to adopt the first response. They practiced the response for the remainder of the 200 trials after adoption. This differs slightly from Maier's procedure in that he has usually required that animals practice the response adopted for a fixed number of trials (160) before entering test training (**5, 6**). Trials were given in sets of 20 massed trials per day, with approximately 20 sec. between the time *S* jumped until he was placed on the stand for the next trial. Any response *S* made which removed him from the jumping stand was considered a trial. If *S* failed to leave the stand within 5 min., *E* removed him and counted it a trial. In cases where escape responses, or waiting beyond the 5-min. criterion became consistent, training was discontinued (*cf.* results).

Cards were alternated from left to right according to the same chance sequences for all groups. For Groups II and III they were left locked or unlocked according to chance sequences which fulfilled the conditions listed in Table 1.

*Test training.*—All rats which continued jumping to the cards were tested to determine whether the response adopted in experimental training was fixated. Rats with position responses received a visual-discrimination problem in which BC was always positive and WC negative. Rats which completed experimental training with card habits (two *S*s from Group III) received a spatial-discrimination problem in which one window (left or right by coin toss) was always positive and the other negative. Two hundred trials were allowed to switch to a new response and *S*s failing to change were considered fixated. The criterion of learning was 30 consecutive, errorless jumps.

## RESULTS AND DISCUSSION

### *Experimental Training*

*Responses adopted.*—Not all rats continued to jump to the windows when punishment was introduced. Three *S*s in the Insoluble-Problem Group and two in the Partial-Reinforcement Group either refused to continue jumping, in spite of the air

blast, or developed some response which removed them from the area of conflict. Of these five *S*s, one dropped off the stand into the net whenever the air came on, three jumped off the stand to the floor, and the fifth consistently had a running fit on the first trial each day and, on the second, waited on the stand beyond the 5-min. criterion. Among the 40 *S*s from all groups which continued to jump to the cards, 38 developed position habits and 2 from the Insoluble-Problem Group developed card habits.

*Number of trials to adopt consistent responses.*—Consistent response is defined as a response which occurs on 30 consecutive trials. The data are presented in Table 2 where it may be seen that the findings confirm learning-theory expectation. Continuous reinforcement leads to quicker acquisition than partial reinforcement, while the nonselective aspect of the Insoluble Problem leads to prolonged variability and slow adoption of a consistent response. All differences between groups are statistically significant at better than the .02 level.

Two *S*s in the Insoluble-Problem Group adopted more than one consistent response during the 200-trial series. They first settled on card habits following initial variability, but gave up the card habits for consistent position habits before the end of the series of trials. Thus, rats in the Insoluble Problem were more variable than those in any other group.

This finding, easily accounted for by learning theory, is in opposition to Maier's claim that insolubility shortens the period of trial-and-error behavior normally shown in learning problems. Maier says:

> . . . the inconsistent administration of punishment in the insoluble problem is recognized by the rat and cuts short the variability in behavior that normally occurs in a problem situation (3, p. 69).

The evidence which Maier has used in arriving at the above conclusion has been taken from an inappropriate comparison. He has compared the number of trials required to learn a visual-discrimination habit under ordinary learning conditions with the number of trials to adopt a position habit in the Insoluble Problem. Since rats generally require many more trials to learn visual discriminations than to learn position habits, the appropriate comparison is between the number of trials required to establish the *same type of response* (position, or visual discrimination) under the two different conditions. The comparisons of the present study fulfill this requirement.

### *Test Training*

The results of test training are shown in Table 3. Only the *S*s which continued to jump to the cards in experimental training received test training. Of these, two from the Con-

TABLE 2

TRIALS TO ADOPT CONSISTENT RESPONSE IN EXPERIMENTAL TRAINING

| Group | *N** | Trials Before First Consistent Response Series | | Trials Before Final Consistent Response Series | |
|---|---|---|---|---|---|
| | | Mean | *SD* | Mean | *SD* |
| I Continuous reinforcement | 15 | 7.4 | 8.06 | 7.4 | 8.06 |
| II Partial reinforcement | 13 | 25.3 | 21.44 | 25.3 | 21.44 |
| III Insoluble problem | 12 | 59.8 | 47.22 | 75.0 | 55.57 |

* Does not include rats which learned escape responses.

TABLE 3

RESULTS OF TEST TRAINING

| Group | *N* Started | *Ss* Continuing to Jump to Windows | | |
|---|---|---|---|---|
| | | *N* | % Fixated | % Showing Variability During Testing |
| I Continuous reinforcement | 15 | 13 | 38% | 62% |
| II Partial reinforcement | 13 | 13 | 92% | 8% |
| III Insoluble problem | 12 | 12 | 58% | 58% |

tinuous-Reinforcement Group learned to jump away from the windows soon after they encountered locked cards for the first time in the discrimination test.

*Effect of partial reinforcement.*—Comparison of the frequency of fixation in the Continuous-Reinforcement Group with that in the Partial-Reinforcement Group reveals the effect of partial reinforcement isolated from that of nondifferential reinforcement. The greater fixation in Group II (corrected $\chi^2 = 6.12$) reveals that partial-reinforcement effects arising from alternating reward and punishment of the correct response in the initial problem are a sufficient condition for the development of a high proportion of fixations. This greater persistence of responses acquired under partial-reinforcement conditions is in line with S–R learning theory and consistent with the findings of every study in the partial-reinforcement literature where resistance to extinction has been investigated (**2**, p. 231).

*Effect of nondifferential reinforcement.*—Comparisons between the Partial-Reinforcement Group and the Insoluble-Problem Group reveal the effect of nondifferential reinforcement with partial reinforcement controlled. The obtained difference in frequency of fixation is not sufficiently reliable (corrected $\chi^2 = 2.31$, $p = .12$), but its direction is opposite to that predicted by Maier's theory which holds that nondifferential reinforcement confuses the *S* and contributes to his frustration.

If we compare Group II and Group III in terms of the number of rats which showed variability at any time in the test problem (regardless of whether they eventually learned), the greater rigidity of responses adopted under the selective-learning conditions of Group II is revealed as significant (corrected $\chi^2 = 5.21$, $p = .02$). This result confirms a conventional S–R prediction that consistent responses adopted under non-selective learning conditions will be less rigid than consistent responses adopted under selective learning conditions when the effects of partial reinforcement are controlled.

*Effect of nondifferential reinforcement and partial reinforcement combined.*—The comparison of frequencies of fixation and learning between the Continuous-Reinforcement Group and the Insoluble-Problem Group, which is Maier's basic experiment, does not reach statistical significance in the present study (corrected $\chi^2 = 0.35$, $p = .50$), although the obtained difference is in the direction predicted by his theory.

*Discussion of group comparisons.*—All of the above differences between groups are easily accounted for by conventional S–R learning theory. This definitely establishes the fact that no new principles are needed to account for Maier's major findings.

In addition, explanation from Maier's point of view appears more difficult. It is unclear from his theory why the Partial-Reinforcement Group should exhibit more fixation than the Insoluble-Problem Group, if nondifferential reinforcement contributes to frustration and fixation. *Ss* in the Partial-Reinforcement Group received slightly more punishment at the cards in the early phases of experimental training (by actual count, 62% vs. 50% before adoption of the first consistent response) and this fact might be used as an explanation. This would imply, however, that the additional early punishment frustrated the *Ss* of Group II. But the fact that all Group II *Ss* which continued to jump to the cards selectively learned the correct response in experimental training argues against the assumption of frustration in these animals at the time they were receiving the greatest proportion of punishment. One of Maier's major points is that responses adopted under frustration are not selectively learned but, instead, are adopted without regard to their consequences.

The failure to find a significant difference in the frequency of fixation between Group I and Group III is not incompatible with S–R learning theory. The fact that Maier has frequently found a larger difference between these groups may be due to the difference in experimental-training procedure described earlier.

*Latencies of correct and incorrect responses in test training.*—One of Maier's reasons for maintaining that fixations are abnormal is that the rat with a position fixation continues to perform the fixated response even after it "knows better"

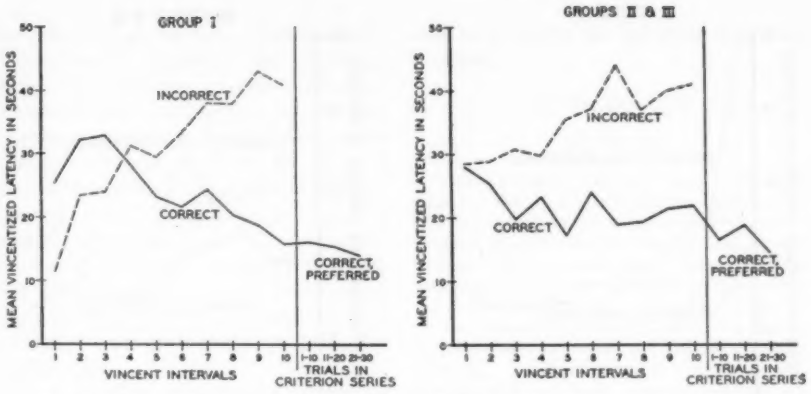


FIG. 1. Curves showing gradual acquisition of differential resistance during discrimination learning among rats which learned to perform the discrimination (Group I, $N = 8$; Groups II and III, $N = 6$)

(3, p. 43). Maier has demonstrated (5) that fixated rats persist in their position habits for many trials after a difference in latency has appeared between jumps made to the positive versus the negative card. In support of the claim that this persistence is abnormal, Maier argues that nonfixated rats change over from position responses to the more adaptive discrimination responses as soon as differential resistance to jumping to the two cards appears.

The requisite data in the Maier studies have never been analyzed in a way which would justify the claim that animals which learn the test problem switch to correct performance *as soon as differential resistance appears*. Such a conclusion can only be drawn from an analysis of the entire course of learning in the test problem.

All *S*s which learned the test problem ($N = 14$) were included in the present analysis. Since only one such rat entered test training with a card habit, the presentation will be simplified by describing the analysis and results as if all rats had been tested on the visual discrimination.

The first analysis was concerned with the latency of jumps with the positive card in the preferred window versus the latency of jumps when the negative card was in the preferred window. The series of trials involved extends from the beginning of discrimination training to the beginning of the criterion series. Since the shape of the learning function is crucial to the interpretations being tested, the latency scores were Vincentized according to Loucks' arithmetical method (1) to express ten equal intervals in the learning process. The use of this method resulted in scores representing average latency per trial per tenth of trials to the beginning of the criterion series. For further comparison the latencies of jumps to the correct card in the preferred window during the criterion series are presented in order to continue the latency function to the final trial.

The results of this analysis are given in Fig. 1. These curves show that latency of jumps to the negative card in the preferred side increases continuously from the beginning of discrimination training, with no indication of a sudden increase in resistance to jumping to the negative card. The initial rise in the curve for latency of jumps to the positive card in the preferred window for Continuous-Reinforcement *S*s probably reflects generalization from the sudden introduction of punishment on the negative card.[5] The curves are clearly predictable from S–R learning theory (*cf.* 9) and fail to support Maier's contention that the normal (i.e., nonfixated) animal shifts from

[5] The difference in latency of jumps to the two cards at the first Vincent interval for Group I *S*s is significant at the 5% level of confidence. Since jumps to both cards had been rewarded on every trial during the long series in which the position habit was practiced prior to test training, it seems likely that the difference is due to some stimulus property of the cards. The *S*s may have preferred the darker card, WC, because, as a stimulus, it was closer on a generalization gradient to the stimulus of an open window than was the other card, BC, given the lighting conditions of the present experiment. In any event, the preference for WC was also shown significantly in experimental training, before the discrimination test began.

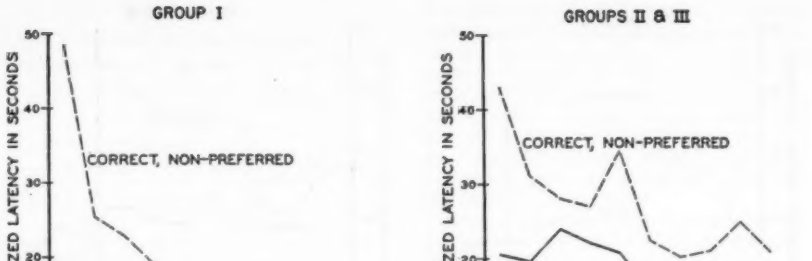


FIG. 2. Curves based on trials after the first break from the old response for rats which learned the discrimination test (Group I, $N = 8$; Groups II and III, $N = 6$)

his position habit to near-correct performance on the visual discrimination as soon as differential resistance appears.

A similar analysis is presented in Fig. 2, in which latency curves are shown for jumps to the correct card in the preferred vs. nonpreferred window. The curves begin with the first jump to the correct card in the nonpreferred window and are Vincentized over the remaining learning to the last criterion trial. As in Fig. 1, the results indicate that learning is a gradual process. Even after the animals break over to the correct card in the nonpreferred window, they must practice jumping to the nonpreferred side for some time before this response is made as quickly as the response to the preferred side. The results do not indicate the insightful kind of break posited by Maier.

Thus, the fact that fixated animals show differential resistance without breaking over to the nonpreferred side cannot be regarded as an abnormality peculiar to their frustrated state. The present findings show that the animals which Maier considers "non-frustrated" also exhibit a differential resistance for some time before the break to a new response, suggesting that the fact some animals perform the discrimination test correctly and others do not represents a quantitative difference in response strength rather than a qualitative difference in the underlying process of response selection.

*The relationship of abortive jumping to the development of fixation.*—The present study also provided an opportunity to analyze the relationship of abortive jumping to the development of fixations. Maier has presented evidence (5) showing that animals which persist in the fixated response during test training tend to adopt two ways of jumping at the cards. In the case of fixated-position responses, the animal jumps straight to the positive card when it appears in the preferred window, but jumps "abortively" when the negative card appears. The abortive jumps can take many forms, but the usual result is that the animal hits the locked card, or the vertical screen to one side of it, with a glancing blow which appears to be relatively unpunishing. He strikes the surface with his feet or the side of his body rather than directly on the nose as in a straight jump.

Maier does not assign abortive jumps a causal role in the development or persistence of fixations. From an S–R learning theory viewpoint, however, the jumps appear to be learned responses of escape from hard bumps

on the nose. These "cushioning" jumps should decrease the effectiveness of punishment at the cards and interfere with the learning of the discrimination test. Stimulus-response theory suggests that if an animal learned to make such jumps in experimental training during the acquisition of his original habit (card or position), he would be much less likely to learn the subsequent test problem than an animal which did not. The following analysis was performed to test this hypothesis.

The number of abortive jumps in experimental training was tabulated for all *Ss* which continued jumping to the cards through test training ($N = 38$). Three types were recorded, each of which had the appearance of being less punishing for the animal than straight jumps. The three types were: (*a*) jumps to one side of either window so that *S* hit the screen at a rather flat angle; (*b*) jumps in which *S* hit the card with his body horizontal and flat, conveniently described as "banking turns"; (*c*) jumps which fell short of the cards.

*Ss* from all experimental groups were dichotomized into Learners ($N = 14$) and Nonlearners ($N = 24$) in terms of their performance in the test problem. The mean number of abortive jumps in experimental training among Learners is 3.9, while the corresponding mean for Nonlearners is 29.2. The Mann-Whitney *U* test reveals that the probability is less than .001 that the two groups come from a common population.

This finding supports the interpretation that abortive jumps represent adaptive solutions which tend to decrease whatever effectiveness punishment at the windows can have in the subsequent test training, and which make it quite unlikely that an animal which has learned an abortive jump will ever learn to perform the test problem.

## Experiment II

### Method

This experiment was run concurrently with Exp. I. The 24 animals which developed position fixations in the first experiment were used as subjects; the same feeding schedule and general routine were maintained, and the same apparatus was used. The air blast was administered according to the standard schedule described in Exp. I.

Two procedures, A and B, were tried for unlearning fixations. Neither involved any guidance (7). *Ss* were randomly assigned to two groups, A–B and B–A. Group A–B was run first in Procedure A and then, if any *Ss* remained fixated, in Procedure B. Group B–A received training by the two methods in the reverse order.

*Procedure A.*—In this method, the animal was forced to jump repeatedly while confronted with the locked negative card in the fixated window and while an open window appeared on the nonfixated side. One hundred trials were allowed, with 20 massed trials per day. If the animal jumped through the open window on his nonpreferred side before the 100 trials were completed, he was given 30 additional practice trials, after which he was tested on the discrimination sequence.

*Procedure B.*—In this procedure, the open window and the negative card were alternated from side to side, as if the open window were the positive card in a regular discrimination sequence. One hundred trials were allowed, with 20 massed trials per day. If the *S* began jumping to the open window on the nonfixated side, he was given 30 trials practice (15 to the open window on each side), after which the positive card was reintroduced in the regular discrimination sequence.

### Results

*Effectiveness of the procedures.*—In Group A–B, 7 *Ss* switched to the open window and 2 learned to climb down from the jumping stand while in Procedure A. One switched to the open window in Procedure B, and 2 failed to change their position habits. In Group B–A, 9 *Ss* began jumping to the open window in Procedure B, 1 switched to the open window in Procedure A, and 2 failed to change. The median number of trials before change

for the 12 *Ss* in Group A–B was 21.0, and the corresponding median in Group B–A was 21.5.

*Performance in the regular discrimination sequence.*—Of the 18 *Ss* which began jumping to the open, nonpreferred window as a result of one or the other of the above training procedures, 13 performed almost perfectly when the discrimination sequence was reintroduced. Five of these animals made no errors, 7 made only one, and 1 made three errors.

The five remaining *Ss* returned to a position habit when tested on the discrimination problem. In three cases the position habit was to the same side as before, but in the case of two *Ss*, which had switched to the open window in Procedure A, the opposite side was fixated, i.e., the side toward which they had just practiced jumping for 30 consecutive trials. Correct discrimination performance was obtained with these *Ss*, however, without the use of any guidance. Since the kind of additional practice which seemed called for varied for different *Ss*, no completely standardized procedure was adopted. Two general rules were followed: (*a*) If the rat showed no differential reaction to the two cards, discrimination training was continued until differential reaction was developed; (*b*) When all animals showed a differential reaction, a graded exposure technique was used. This procedure involved reintroducing the open window on the nonpreferred side, followed by a gradual reintroduction of the positive card with the window partially open.

The significance of the results in Exp. II is that 82% of the fixated rats abandoned their position fixations when the learning conditions were modified to make the application of differential reward more effective. As in the results of Exp. I, this suggests that the fixations are not in any way symptomatic of an abnormal condition. Rather, they are strong habits which can be unlearned, if the conditions are made appropriate.

## Summary and Conclusions

Two experiments are reported. The purpose of the first was to isolate the effects of partial reinforcement and nondifferential reinforcement in a situation similar to the one from which Maier has drawn his evidence for "abnormal fixations." The same apparatus and general method as used by Maier were employed. The design involved training three groups of rats in an initial habit under different experimental training conditions. Group I, the Continuous-Reinforcement Group, was trained in a position habit under conditions of invariable reward for the correct response and invariable punishment for the incorrect. Group II, the Partial-Reinforcement Group, learned a position habit where the correct response was rewarded half the time and punished the other half, while the incorrect response was punished each time it occurred. Group III, the Insoluble-Problem Group, received training in which any card or position habit adopted resulted in reward on half the trials and punishment on the other half. Following experimental training, all animals were tested for their ability to learn a new response which required abandonment of the one learned under experimental conditions.

The purpose of the second experiment was to demonstrate techniques other than manual guidance for teaching rats to abandon position fixations.

The major results and conclusions from both experiments may be summarized as follows:

1. In the initial training condition, rats in the Insoluble-Problem Group

required significantly more trials to establish a consistent response than rats in any other group. It is concluded that nondifferential reinforcement per se increases variability, as would be expected from a learning point of view, and that Maier's conclusion to the contrary is based on an inappropriate comparison.

2. Lack of differential reinforcement is not necessary to obtain a high proportion of fixations in this situation, but partial reinforcement of the first habit developed is a learning condition of primary importance in the development of fixations.

3. Responses adopted under partial-reinforcement conditions in a selective-learning situation were found to be more rigid in test training than those adopted in the Insoluble Problem.

4. Analysis of latency data revealed that the learning of the test-discrimination habit is continuous, with differences in latency between correct jumps and incorrect jumps appearing in the normal course of learning long before the animals make correct choices. Since differential resistance to the positive and negative cues is shown to a striking degree in animals which eventually learn the test, the analogous behavior shown by fixated rats cannot be taken as an indication of abnormality.

5. Analysis of abortive jumps revealed that animals which learn abortive jumps during acquisition of the initial response in experimental training are quite unlikely to learn the subsequent test discrimination. Abortive jumps may be interpreted as being learned on the basis of reduction in punishment. By reducing punishment, they interfere with the learning of the correct response.

6. Results of the second experiment indicate that fixations may be unlearned by applying relatively simple techniques based on learning principles, and without the use of any guidance.

7. Conventional S–R learning theory can account for behavior which Maier considers incapable of being explained without a new set of principles.

(Received April 5, 1952)

## References


1. Hilgard, E. R. A summary and evaluation of alternative procedures for the construction of Vincent curves. *Psychol. Bull.*, 1938, **35**, 282–297.
2. Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
3. Maier, N. R. F. *Frustration, the study of behavior without a goal.* New York: McGraw-Hill, 1949.
4. Maier, N. R. F., & Ellen, P. Can the anxiety-reduction theory explain abnormal fixations? *Psychol. Rev.*, 1951, **58**, 435–445.
5. Maier, N. R. F., Glaser, N. M., & Klee, J. B. Studies of abnormal behavior in the rat: III. The development of behavior fixation through frustration. *J. exp. Psychol.*, 1940, **26**, 521–546.
6. Maier, N. R. F., & Klee, J. B. Studies of abnormal behavior in the rat: XII. The pattern of punishment and its relation to abnormal fixations. *J. exp. Psychol.*, 1943, **32**, 377–398.
7. Maier, N. R. F., & Klee, J. B. Studies of abnormal behavior in the rat: XVII. Guidance versus trial and error in the alteration of habits and fixations. *J. Psychol.*, 1945, **19**, 133–163.
8. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.
9. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
10. Thorndike, E. L. *The fundamentals of learning.* New York: Teachers College, 1932.

# AN EXPERIMENTAL DEMONSTRATION OF UNCONSCIOUS MEDIATED ASSOCIATION

B. R. BUGELSKI AND D. P. SCHARLOCK

*University of Buffalo*

Some years ago McGeoch noted that "the theory of the existence of mediated associations is an old one which has generated more discussion than straightforward experimental study" (**5**, p. 94). In his recent revision of McGeoch's text, Irion (**6**) saw no good reason for altering the statement. The concept of mediation is of great potential value for the psychological analysis of learning, thinking, and insight as Hebb (**2**) has recently demonstrated, but unequivocal demonstration of such mediation on a verbal level is as yet lacking. There have been a number of studies of mediated association on a nonverbal level (**1**, **4**, **8**), but in these studies there is some question as to whether the mediation is a function of sensory-sensory association, stimulus generalization, or whether an intermediate response occurs and functions as a mediator. Hilgard and Marquis (**3**) note that in Lumsdaine's experiment the intervening eyelid response was not always present in test trials even though the response that had never been associated with the test stimulus did occur. The recent report of Wickens and Briggs (**9**) purports to demonstrate the need for a common response to two different stimuli as a basis for mediation. Their demonstration is not free from the criticism that the *Ss* may have operated under different self-instructed sets in the control experiments and, therefore, did not respond to the nonconditioned stimulus because they had discriminated the relevant and irrelevant stimuli during the preliminary training.

Peters' (**7**) report still remains as the only demonstration of mediation in verbal learning that Irion cites. It is important to note that Peters found such association for only a few *Ss*, in only two out of nine experiments, and only where *Ss* were able to make use of the common item "perceptually or ideationally present at the time of recall." When the common item is ideationally present, there is some question as to whether the process can be fitted into the concept of mediation as Hebb uses it: "Two concepts may acquire a latent 'association' without ever having occurred together in the subject's past experience" (**2**, p. 132). Certainly a broader potential application of the mediation hypothesis can be developed if such mediation operates without awareness on the part of the *Ss*. If, as in Peters' experiments, *S* specifically makes use of two previous sets of associations by which to arrive at a third, he is using something more akin to a mnemonic device than to mediated association. It should be clear to the reader that the mediation hypothesis involves two separate questions: (*a*) Does the mediation depend on the occurrence of the intermediate response as an agent for eliciting the new response? and (*b*) Must *S* be aware of such occurrence if it does take place?

Hebb (**2**) implies and Lumsdaine (**4**) indicates that the mediating response need not actually occur. Peters' conclusion, just cited, indicates that in verbal material, at least, awareness is necessary. The experiment to be described here deals with

the second of the two questions. It is not expected to shed light on the answer to the first question.

## Method

*Experimental design.*—The general outline of the experiment followed the pattern of Peters' nine experiments. The *Ss* first learned one set of paired-associate nonsense syllables so arranged that the stimulus and response syllables can be labeled A and B. They then learned another set of paired syllables with the original response syllables serving as stimuli so that the sequence B–C was established. The last stage consisted of a test for mediation in requiring *Ss* to learn the A–C pairs. Assuming that mediation has a chance to operate under these circumstances, *Ss* should learn the A–C list with comparative ease, all other factors considered equal or controlled. Because the A–C list could actually be learned more or less easily for various reasons irrelevant to the experimental hypothesis, e.g., practice, proactive inhibition, familiarity with the syllables, or even a chance selection of easy- or hard-to-learn associates in the A–C list, it was necessary to make each *S* serve as his own control and learn in his A–C list a number of paired associates which had been learned in the A–B, B–C sequence and an equal number where such an arrangement did not hold. To achieve this effect, half of the A–C list was constructed directly from the previously associated pairs, thus A–C units were made up of the A syllable and the C syllable that had been associated with a common B syllable. These A–C's will hereafter be called the experimental syllables. The other half of the A–C list was composed of the remainder of the A syllables paired at random with the remainder of the C syllables, so that no A–C pair in this half of the list had a common and specific B-syllable associate. These pairs will hereafter be called the control syllables. The arrangement of the experimental and control syllables can be appreciated from the paradigm below:

| First List | Second List | Third List | |
|---|---|---|---|
| A1 — B1 | B1 — C1 | A1 — C1 | Exper. |
| A2 — B2 | B2 — C2 | A2 — C2 | |
| A3 — B3 | B3 — C3 | A3 — C4 | Control |
| A4 — B4 | B4 — C4 | A4 — C3 | |

From the arrangement above it is obvious that if an associative inhibition is implied, the *Ss* can be expected, other things equal, to learn the experimental syllables better either because of high positive transfer from the common elements or because of associated interference operating to reduce learning for the control pairs. A high negative correlation, then, is to be expected between the learning scores for the two sets of materials.

The design, as thus far described, however, does not escape the difficulty that either of the two halves of the A–C list may, by chance, be more difficult than the other half. If the randomized half were learned more easily, it might be interpreted to mean that no mediation had been operating. To overcome this factor, half the *Ss* learned the pairs with the roles of the syllables reversed so that the formerly randomized or "control" syllables were now used to build up lists of A–B and B–C where the B was a common item, and the experimental A–C's which were previously associated with a common B were now randomized in the A–B and B–C lists. The experimental and control lists, then, consisted of exactly the same syllable pairs. They differed only in their history.

*Apparatus.*—Syllables for constructing the paired associates were taken from Glaze's lists of syllables with 40–50% association value. Six lists, each consisting of 16 paired associates, were prepared to fit the needs of the design just described. Thus there were A–B, B–C, and A–C lists for one group of *Ss* wherein the experimental syllables or first eight pairs of the A–C list were composed of A and C syllables with a common B associate and eight where the B associate was randomized. The other three lists were arranged so that in the A–C list the original experimental syllables now had their B associate randomized and served as control syllables while the former control syllables were systematized (associated with a common B) and served as experimental syllables. A second group of *Ss* learned the latter three lists. This arrangement permits the addition of data for both groups of *Ss* and for both groups of experimental and control syllables. It should be noted, however, that half of the *Ss* learned half of the experimental syllables and half of the control syllables, the remaining half of the *Ss* learned the other halves of the experimental and control syllables. The lists of A, B, and C syllables appear in Table 2. From this table the general arrangement of the A–B, B–C, and A–C lists can be inferred.

The syllables were block printed on 3 × 5-in. white cards and were viewed through an opening in a screen. Behind the screen a motor-driven exposure device consisting of a slotted screen was arranged to expose the stimulus half of each pair of syllables for 3 sec. and the response half for an additional 3 sec., with 4 sec. between pairs.

*Procedure.*—Each *S* learned three lists of paired associates; one at each of three separate learning sessions. The *Ss* learned to anticipate each correct response to a criterion of five perfect trials. As the criterion for each pair was met, that pair was eliminated to prevent differential

overlearning. The cards were shuffled between trials to prevent serial learning. Learning was usually completed in 1 hr.; 48 hr. after learning the first list, *Ss* learned the second list, and 48 hr. later, the third list.

*Subjects.*—Twenty senior students in psychology, all of whom had nonsense syllable learning experience, served as *Ss*. Each *S* served as his own control. Half of the 20 *Ss* learned the original set of materials, i.e., the first three lists, and the other ten learned the same syllables arranged to control the factor of difficulty as described above.

## Results

Because of the need for controlling the possible differential difficulty of the experimental and control syllables, the results for the two groups of ten *Ss* have been combined and treated as one lot. This procedure is justified only if the two groups are not significantly different in learning ability. The mean number of trials to learn the 16 paired associates for the two groups of *Ss* was 16.9 and 17.2, with *SD*'s of 8.3 and 4.9, indicating essential equality of the two groups of *Ss*.

Because of individual differences among *Ss* and because of unusual difficulties some *Ss* experienced with some pairs, no significant differences could be established between group results for either median or mean number of trials to learn the experimental as opposed to the control materials, although in each case the differences favored the experimental syllables. Examination of the order of learning the two sets, however, suggested that a significant difference between the two groups of syllables could be established. Accordingly, the mean rank order of learning of the experimental syllables was compared with the mean rank of the controls (see Table 1). Using a single-tail *t* test and taking account of the high negative correlation ($-.87 \pm .06$) in the calculation of $SD_{diff.}$, the difference between the mean rank orders is 1.2 which is significant between the .02 and .05 levels. It should be noted that 12 of the 20 *Ss* learned the experimental syllables before they learned the controls (in terms of mean rank), 2 *Ss* averaged the same, and 6 *Ss* had a lower mean rank for the control syllables. When the syllable pairs themselves are examined, the same general picture is revealed. It will be recalled from the experimental design that all 16 pairs of syllables served as experimental and control syllables (eight syllables for each group of ten subjects). When a comparison is made between the median number of trials to learn the 16 syllable pairs on their control function (see Table 2), it is found that 12 of the 16 pairs are learned more rapidly in the experimental arrangement. The mean difference between the medians is 1.7 which is significant between the .005 to .01 levels as determined by using a single-tail *t* test.

One additional finding of importance is that because of the combination of factors in the learning situation (large numbers of syllables to learn, 48-hr. periods between sets, and probably most important, the mixture for each *S* of experimental and control syllables so that no logical or mnemonic system could work consistently) none of the *Ss* was able to report any correct appreciation of the nature of the experiment and most assuredly did not verbalize a pattern of A–B, B–C, A–C in the learning of the third list.

TABLE 1

Mean Rank Order of Learning Experimental and Control Syllables by 20 *Ss*

| Syllables | Mean | *SD* | $\sigma_m$ | $r$ exper. and control |
|---|---|---|---|---|
| Exper. | 7.9 | 1.64 | .38 | $-.869 \pm .056$ |
| Control | 9.1 | 1.57 | .36 | |

As far as can be judged by results of interrogation of the *Ss*, they learned the experimental A–C syllables more efficiently than they did the controls without the benefit of mediations that were "perceptually or ideationally present." None of the *Ss* consciously resorted to the use of the mediating B syllable in their attempts to learn the A–C list.

## Discussion

The results presented appear to form a reasonably clear-cut demonstration of mediated association in the learning of verbal material. The fact that *Ss* did not consciously use the intervening common item in the more efficient learning of the experimental materials suggests that such conscious verbal interpolation is unnecessary and that it is sufficient to have had the experience to make use of it in a different connection. Whether reactions involving B syllables occurred unconsciously or at all cannot be determined from this experiment. We have, then, a sample of the type of evidence Hebb (2) alludes to in his automatic association of cell assemblies and phase sequences. The gratuitous learning which is assumed to have facilitated the learning of the experimental material attests to the significance of mediated association in the acquisition of new learned patterns.

Because the present findings reveal a significant amount of mediated association and Peters failed to find significant evidence of such mediation, some possible reasons for the differences might be considered. In the present study a rather high criterion was employed to insure adequate learning of the prospective mediation material. Peters generally used a criterion of two correct anticipations and this might not have been high enough. Peters also generally used an anticipation test instead of a learning test. It may be that he was using too difficult a test situation, considering his low learning criteria and the nature of his materials. A last comment concerns the order of presentation of the learning material. In the present study the order of learning the successive lists was A–B, B–C, A–C. Peters frequently used other orders such as A–B, C–B, A–C. It may well be that the reversal of the B–C may prove a more difficult condition for future mediation as it corresponds somewhat to backward conditioning which is generally recognized to be difficult to establish.

TABLE 2

The Median Number of Trials to Learn Each A–C Pair when the Same Pairs Are Learned as Experimental and Control Materials

| | Syllables | | | Exper. Median | Control Median |
|---|---|---|---|---|---|
| | A | B | C | | |
| 1 | Fuj | Mup | Qeh | 6.5 (y) | 7.0 (x) |
| 2 | Gud | Sif | Nal | 4.0 | 7.0 |
| 3 | Jek | Fah | Tik | 6.5 | 6.0 |
| 4 | Lud | Tek | Lix | 1.5 | 6.0 |
| 5 | Moh | Yuk | Reh | 7.5 | 4.5 |
| 6 | Pah | Baz | Tij | 1.5 | 5.0 |
| 7 | Viq | Zes | Cax | 4.0 | 6.0 |
| 8 | Weg | Vag | Puv | 3.0 | 10.5 |
| 9 | Cey | Hix | Kiv | 5.0 | 8.5 |
| 10 | Dak | Woh | Gew | 3.0 | 3.5 |
| 11 | Hax | Kor | Yob | 8.5 | 6.0 |
| 12 | Kug | Jep | Zan | 8.5 | 7.5 |
| 13 | Nof | Dal | Huy | 6.5 | 9.5 |
| 14 | Sux | Gow | Mog | 5.0 | 8.0 |
| 15 | Yob | Qur | Jax | 7.5 | 9.0 |
| 16 | Zew | Lah | Bup | *7.0* | *8.5* |
| *Mean* | | | | 5.3 | 7.0 |
| *SD* | | | | 2.3 | 1.9 |
| $SD_m$ | | | | 0.59 | 0.49 |

## Summary

1. Twenty college seniors learned three lists of paired associates so arranged as to provide a possibility of mediated association. The three lists followed the order of A–B, B–C, A–C.

2. The A–C syllables were learned significantly better than control syllables which were the exact duplicate of the experimental syllables except for different prior experience of the *Ss*.

3. The *Ss* benefitted from prior learning of the A–B, B–C lists without reporting any perceptual or ideational use of the material.

(Received April 7, 1952)

## References

1. Brogden, W. J. Sensory preconditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.
2. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
3. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
4. Lumsdaine, A. A. Conditioned eyelid responses as mediating generalized conditioned finger withdrawal reactions. *Psychol. Bull.*, 1939, **36**, 650.
5. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, 1942.
6. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* (2nd Ed.) New York: Longmans, 1951.
7. Peters, H. N. Mediate association. *J. exp. Psychol.*, 1935, **18**, 20–48.
8. Shipley, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12**, 337–357.
9. Wickens, D. D., & Briggs, G. E. Mediated stimulus generalization in sensory preconditioning. *J. exp. Psychol.*, 1951, **42**, 197–200.

# GENERALIZATION OF EXTINCTION OF AN INSTRUMENTAL RESPONSE TO STIMULI VARYING IN THE SIZE DIMENSION[1]

J. W. KLING

*Brown University*

The phenomenon of generalization has been observed to occur under a variety of conditions since it was first reported by Pavlov (17). Results obtained from the Pavlovian studies, and later evidence provided by Anrep (1), Bass and Hull (2), and Hovland (10, 11) indicate that the effects of extinction as well as the effects of reinforcement can be expected to generalize to novel stimulus situations. Utilizing these data, Hull (12) has assumed similarity of shape for the inhibitory and excitatory generalization gradients. Likewise, Spence (20) had earlier made a similar assumption in demonstrating the means by which transposition phenomena could be predicted in the discrimination situation.

Until recently, there has been little evidence concerning the possible generalization of extinction in an instrumental-response situation. Youtz (26) and Ellson (5) have shown that extinction of one bar-pressing habit would hasten the extinction of a second such habit, but the functional relation between variations in one of the stimulus dimensions and the amount of generalization was not investigated. More recently, Thompson (23) reported the results of a study in which albino rats were trained to approach four sizes of visual stimuli and then extinguished on one of the extremes of the size dimension. Following attainment of the criterion of extinction, each animal was given multiple testing on the stimuli to ascertain the effects of extinction on one extreme of the size continuum. Thompson found significant amounts of generalized extinction, and very steeply falling gradients.

The extent to which these results indicate the primary generalization gradients for inhibitory tendencies is not known. What effect the repeated testing of the animals had upon Thompson's obtained gradients cannot be determined, although it may be assumed that much of the transitory effects of extinction were allowed to dissipate by the 30-min. interval which was provided between the several test trials.

What would seem to be a more serious methodological difference exists in the early studies of the phenomenon. Anrep, Bass and Hull, and Hovland all utilized a differential reinforcement procedure which could be expected, according to Hull's theoretical analysis (13), to result in a "narrowing" of the gradients. Presumably such gradients reflect the interaction of generalized excitation and extinction, and should therefore be logically termed "discrimination gradients" (15). This effect was discussed by Grice and Salz (9) in connection with their study of the generalization of excitation effects in an instrumental-response situation. By testing each animal on one stimulus

[1] This paper represents a major portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology of the University of Illinois in 1951. The writer wishes to express his appreciation to Professor G. Robert Grice for his invaluable aid in the direction of the investigation; and to Professor Arthur L. Irion for his many helpful criticisms during the preparation of the original thesis.

only, they obtained generalization gradients which were not complicated by the effects of nonreinforcement of other stimuli.

The present study was designed to afford an estimate of the extinction-generalization gradients for the visual size dimension of the albino rat. Each *S* was trained to approach a pair of stimuli presented one at a time in a predetermined order. One of the stimuli was then presented without reinforcement until the criterion of extinction was attained. At this point the second stimulus of the training pair was presented. The latency of response in this test situation was utilized as a measure of the extent to which extinction of the one habit had generalized to the other.

## Procedure

*Subjects.*—The *S*s were 112 experimentally naive albino rats of the strain maintained at the University of Illinois Animal Psychology Laboratory. There were 53 males and 59 females, divided approximately equally among the experimental groups. The ages at the start of training ranged from 66 to 112 days, with the median ages for the groups being approximately equal. Twenty-four *S*s were discarded prior to the extinction training: 7 for illness, 6 for failure to eat during preliminary training, 6 for inability to learn the door-opening habit, and 5 for excessive delays in running during the acquisition trials. After all discards there remained a total of 88 rats, 11 in each group. The *S*s were run under approximately 23-hr. hunger drive, being fed the remainder of the daily diet in individual feeding cages after each had completed the daily runs.

*Apparatus.*—The stimulus objects used in the experiment were white discs devised by Grice (7). They were cut from sheet metal and enameled white. Each disc had in its center a small door 2.8 cm. square which was hinged from behind and set flush with the surface of the disc. The small doors could be opened by the rat's nose to allow access to a food dish located on the back side of the disc. All doors were equidistant from the floor so that the responses required of *S*s were the same regardless of the area of the disc.

Disc areas were chosen which represented approximately equal log steps: 20, 32, 50, and 79 sq. cm. These stimulus objects were presented in a Grice Successive-presentation Discrimination Apparatus (8). Briefly, this consisted of a 24-in. alley containing vertically sliding doors at either end and the center. A holder for the discs which could be slid against the end of this alley was arranged so that the discs could be changed with very little loss of time between trials. The 24-in. alley had a pivot in the center of its floor which allowed it to be reversed at the end of a trial and thus place *S* in position for the start of the following trial without handling. The disc holder was constructed so that it formed what was essentially a 2-in. extension of the alley when it was slid into position.

All portions of the apparatus except for the white discs were painted a flat black. The top of the alley was of ordinary window glass, while the top of the disc holder was a panel of black masonite. Illumination was provided by a single 40-w. bulb suspended 20 in. above the center of the starting end of the alley. The walls and ceiling of the experimental room were painted black to reduce extraneous light in the discrimination apparatus.

To the human observer, the white discs appeared to stand out from a uniform black background, and the small reaction door in the center of the disc was scarcely noticeable at distances greater than 6 ft.

A preliminary investigation indicated that the nonreinforced trials used to achieve extinction might result in the entire habit of running in the apparatus being extinguished. To prevent this, some means of rewarding the general habit of running through the center door into the reaction end of the alley had to be devised. For this purpose, a black square 10 cm. on an edge and equipped with a small reaction door in its center was constructed. When in place in the disc holder, this square appeared to the human observer to be practically nondistinguishable from the usual black background provided for the white discs. Responses to the black square were always reinforced, a procedure which was successful in maintaining the running responses during the extinction trials.

*Preliminary training.*—The conditions of the experiment are given in Table 1. All *S*s were placed on one feeding per day for at least seven days before beginning training. Adaptation to the apparatus consisted of placing *S*, under approximately 23-hr. hunger drive, in the alley with the center door lowered. The food dish located on the back side of the white disc was loaded with ten cylindrical pellets made from a mixture of Purina Laboratory Chow and white flour, weighing approximately .06 gm. each. The small response door in the center of the disc was fastened open, the center door of the alley was

raised, and *S* was allowed to enter the reaction end of the alley and eat the ten pellets. The vertically sliding door in front of the disc was then lowered, the alley reversed, the food dish reloaded with ten more pellets, and the procedure repeated. Any *S* not consuming the first ten pellets within 45 min. was discarded. No time limit was set upon acceptance of the second ten pellets.

Day 2 consisted of ten trials during which the response door was gradually closed so that by Trials 9 and 10, it was completely closed and *S* was opening it without any assistance from *E*. Any *S* not acquiring this habit at this time was discarded.

During all preliminary training, the area of disc employed was the same as that on which the particular *S* was to be extinguished (see Table 1).

*Acquisition.*—Days 3, 4, and 5 were devoted to the acquisition of the running habit in response to the white discs and the black square. Each *S* was given ten reinforced runs to the disc on which he was to be extinguished, ten reinforced runs to the disc which was to be used for generalization testing, and five reinforced runs to the black square. The order of presentation of the extinction (E) and the generalization test (T) stimuli was the same for all *S*s: ETEETETTE TETTEETETTE. The five runs to the black square were interspersed at the convenience of *E*.

Each successful door-opening response was rewarded with one of the .06-gm. pellets. Latency of response was recorded for every trial with a hand-operated stop watch. The watch was started when the center alley door was rising; the watch was stopped when *S*'s nose moved the response door in the center of the disc. The center door was raised when *E* was ready, and regardless of the position of *S* in the starting compartment. The *S*s learned to take up a position close to the center door and would usually be through the door before it was completely raised. Movement of the response door by *S*'s nose was arbitrarily judged by *E*. Latencies were recorded to the nearest tenth of a second. Any *S* which failed to respond in 10 min. on any trial was eliminated from the study.

*Extinction.*—Day 6 began extinction training. The first four trials of this day were reinforced, the order of presentation of the test and extinction stimuli being: TETE. Immediately following the response on Trial 4, the extinction disc was reloaded and the small response door in the center of the disc was locked from behind so that it would open only ½ in. The extinction trials were then begun.

During extinction trials, each *S* was presented with his extinction disc and the black square. The response doors of the extinction discs were always locked and the food trays loaded (hence, responses to these stimuli were never food-rewarded). Responses made to the black square were always rewarded. Interspersal of the black square proved an effective means of maintaining the running habit: The animals would crouch by the center door and start through as soon as the upward movement of the door began. The order in which the extinction (E) and the black (B) stimulus objects were presented was: EBEEBEBBEBEBBEEBEBBEEBEEBEBBEB.

Each extinction day consisted of 15 reinforced trials to the black square and 15 unreinforced trials to the white extinction disc. The criterion of extinction was no response on 4 out of 5 successive trials to the white disc, with "no response" arbitrarily defined as 30 sec. without moving the small response door in the center of the disc. Latency of response was recorded for each trial. Immediately upon achieving the criterion of extinction, *S* was removed from the apparatus and placed in an individual feeding cage for the receipt of the remainder of the daily diet.

Every effort was made to eliminate sudden or noxious stimuli from the procedure. The attempt was made to produce extinction through nonreinforcement of the instrumental response without the additional complications attendant upon the use of "punishment."

*Generalization testing.*—On the day following this attainment of the extinction criterion, *S* was again run under the same conditions until the identical criterion of extinction was reached once more. At this point an immediate shift was made to the test trials. The test disc was presented with a pellet in the food tray and the small response door unlocked. The black square was not presented during the test trials.

Latency of response was recorded. To simulate as closely as possible the conditions utilized during the extinction trials, in any case in which

TABLE 1

EXPERIMENTAL CONDITIONS

| Group | Area of Circles Employed* | | | |
|---|---|---|---|---|
| | Pre-training | Acquisition | Extinction | Testing |
| 79–79 | 79 | 79 & 79 | 79 | 79 |
| 79–50 | 79 | 79 & 50 | 79 | 50 |
| 79–32 | 79 | 79 & 32 | 79 | 32 |
| 79–20 | 79 | 79 & 20 | 79 | 20 |
| 20–20 | 20 | 20 & 20 | 20 | 20 |
| 20–32 | 20 | 20 & 32 | 20 | 32 |
| 20–50 | 20 | 20 & 50 | 20 | 50 |
| 20–79 | 20 | 20 & 79 | 20 | 79 |

*Where the same size circle is employed for both extinction and generalization testing, two separate circles of identical area are employed.

TABLE 2

LATENCY IN SECONDS TO FIRST RESPONSE ON THE TEST SERIES*

| Extinction Stimulus | 79 sq. cm. | | | | 20 sq. cm. | | | |
|---|---|---|---|---|---|---|---|---|
| Test Stimulus | 79 | 50 | 32 | 20 | 20 | 32 | 50 | 79 |
| | 63.5 | 101.4 | 4.2 | 2.7 | 34.3 | 6.2 | 5.2 | 12.0 |
| | 15.3 | 2.9 | 14.8 | 1.6 | 59.5 | 16.5 | 135.2 | 5.3 |
| | 67.8 | 239.6 | 2.7 | 1.8 | 23.0 | 40.8 | 95.7 | 2.0 |
| | 6.2 | 9.7 | 12.3 | 2.7 | 21.7 | 8.5 | 5.5 | 35.8 |
| | 13.6 | 8.1 | 37.0 | 5.0 | 128.5 | 3.4 | 5.9 | 14.7 |
| | 33.4 | 31.8 | 2.1 | 3.9 | 136.0 | 14.3 | 3.9 | 1.6 |
| | 12.7 | 94.9 | 8.1 | 1.7 | 14.7 | 4.1 | 2.5 | 3.1 |
| | 542.9 | 4.9 | 2.5 | 31.8 | 206.4 | 12.5 | 5.6 | 3.6 |
| | 19.3 | 35.8 | 9.4 | 1.2 | 26.9 | 17.9 | 91.5 | 3.5 |
| | 26.8 | 45.0 | 3.8 | 1.5 | 37.3 | 25.1 | 3.5 | 2.4 |
| | 29.7 | 69.1 | 22.9 | 4.0 | 50.7 | 28.3 | 35.5 | 2.8 |
| Median | 26.8 | 35.8 | 8.1 | 2.7 | 37.3 | 14.3 | 5.6 | 3.5 |

* Each value in the body of the table represents one animal.

*S* did not respond within 30 sec. the end door was lowered, the alley reversed, and the disc again presented. This procedure was repeated until *S* responded, and the latencies of the several trials were then summed to provide the latency value for the first response. Each *S* was run until a test trial latency equal to, or smaller than, the median latency of his fastest block of five trials during any of the three acquisition days was obtained. In all cases, at least ten test trials were run.

The use of the disc holder made possible the rapid switch from one stimulus object to another, and the time between the last extinction trial and the first test trial was thus roughly the same as the intertrial interval during all phases of acquisition and extinction. Elapsed time between *S*'s response to the stimulus object and the start of the next trial was approximately 11 sec.

For the two groups tested on the same disc area as was used during extinction (i.e., Groups 20-20 and 79-79), the testing utilized a new disc of the same size. By this procedure, it was hoped that cues other than area of disc would be varied as much for these groups as for the other groups.

## RESULTS

The latencies for all *S*s are presented in Table 2. Inspection of these values will reveal the positive skewness existing within each of the groups. Transformations of these scores (3) failed to provide measures which would satisfy the assumptions of analysis of variance concerning the population parameters. Under these conditions, the "best" representation of the group central tendencies is afforded by the medians.[2] Accordingly, Fig. 1 presents the median latency for each group as a function of the area, in approximate log steps, of the test stimulus.

The median latency on the last five trials in which the black square was presented during extinction training was employed as a measure of the speed of response for each *S* in the apparatus. Comparisons of the groups by means of Wilcoxon's (24) method revealed no significant differences, i.e., the conditions of training did not result in characteristically different rates of response in the apparatus for the different groups. Since the test trials followed these runs to the black square without *S* being handled, it seems plausible to assume that any significant group differences found on the test trials are somehow related to

[2] A plot of the data according to the procedure suggested by Schlosberg and Stanley (19) was carried out. The derived values did not differ essentially from the group medians and hence are not included in the report.

the different disc areas utilized as test stimuli.

It is also possible that any test-trial differences might somehow be related to the fact that extinction was to a performance criterion, resulting in variations in the number of trials of extinction each *S* received. The number of extinction trials for each *S*, when converted to logarithmic values, yielded a distribution which was essentially normal and which satisfied Bartlett's test of homogeneity of variance (**4**). Analysis of variance revealed that the estimated variance based upon the within-groups sum of squares was larger than that based upon the between-groups sum of squares. These results suggest that the number of extinction trials probably is not significantly related to the group differences found on the test trials.

All groups showed significant degrees of generalized extinction when latency of response in the test series was compared with the estimate of running speed for each *S* based on the median score of the last five runs to the black square. These values were compared using Case II, suggested by Wilcoxon (**24**). In each group, then, the median latency of response on the test trials was significantly greater than the latency under conditions of comparable deprivation and comparable acquaintance with the apparatus and procedure.

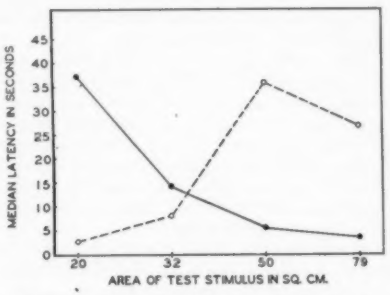


FIG. 1. Generalization functions, showing median latency to first response for each group on the test trials. Solid line represents generalization to larger areas; broken line, generalization to smaller areas.

TABLE 3

SIGNIFICANCE OF DIFFERENCES BETWEEN ADJACENT GROUPS

| Compared Groups | $p$ | Compared Groups | $p$ |
|---|---|---|---|
| 20–20 & 20–32 | .003 | 79–79 & 79–50 | not sign. |
| 20–32 & 20–50 | not sign. | 79–50 & 79–32 | .024 |
| 20–50 & 20–79 | .053 | 79–32 & 79–20 | .028 |

The test scores of adjacent groups were compared and differences estimated according to several conventional nonparametric tests (**6, 16, 24**). There was, as should be expected, agreement among the tests in every instance. The results given in Table 3 are based upon the Mann-Whitney *U* test (**16**). It will be noted that the general downward trend of the gradients is confirmed by these analyses, while the inversion between groups 79–79 and 79–50 is not statistically significant.[3] The extreme conditions (with respect to the independent variable) might be expected to result in the greatest differences in latencies. Comparisons of 79–20 with 79–79, and 20–79 with 20–20 substantiated this assumption. By the Mann–Whitney *U* test, the differences obtained were, in both cases, beyond $p = .0002$ (for one tail of the distribution).

Conversion of the acquisition trial latencies to logarithms provided scores which satisfied the criteria for the use of the *t* test. Comparisons were made of the latencies to the two discs used

[3] No procedure whereby these data could be analyzed in terms of the forms of the functions has been found. Under these conditions, the hypothesis that both gradients are merely chance deviations from some common curve cannot be tested.

for each group. In no case was a significant difference in latency of response obtained, suggesting that area of disc was not a significant factor in determining latency of response during the three days of the acquisition training.

## Discussion

Gradients of generalization of extinction have been presented which represent the behavior of white rats in a visual size-discrimination problem. Inspection of Fig. 1 will reveal that the general falling nature of gradients frequently hypothesized for generalization has been obtained in the present study. The actual shape of the gradients will be seen to be nonsymmetrical. In this aspect, there is a marked departure from the gradients utilized by Spence (**20, 21, 22**) in his discrimination theory. There is one point which is of extreme importance in interpreting this variance: The gradients assumed by Spence were chosen for their ability to predict discrimination and transposition, and as such were not hypothesized as representing the "true" form of generalization gradients. Indeed, Spence has upon occasion introduced variations in the assumed functions (**21,22**), a procedure which should emphasize the essentially heuristic nature of the functions.

In other instances, notably in Hull's (**12**) system, it has seemed necessary to make certain assumptions concerning the nature of the generalization gradients per se. Supposedly such theoretical functions represent generalization with all other factors held constant. Seen in this relation, empirically determined gradients would represent the "true" gradient as influenced by certain "distorting" factors. One of these would almost certainly be the influence upon behavior of variability in the intensity of the stimuli presented. Reaction-time studies have demonstrated quite conclusively that RT is some decreasing function of stimulus intensity (**25**), and the significance of this relation for behavior theory has been discussed by Hull (**14**).

If, in the present study, we are willing to assume that area variations have an effect similar to that of intensity changes, then we would expect that the latency of response to a small area would be greater than its "true" value, while the latency of responding to a stimulus of large area should be smaller than its expected value. If it were possible to eliminate this stimulus factor from the results, we might possibly find our gradients displaying the symmetry postulated by Hull (**12**).

This interpretation of our test-trial results suffers when we recall that an analysis of the latencies obtained during acquisition did not show significant differences for area variation. Very possibly, this is due to the training given the *Ss*: For each group, two days of preliminary training were given with the disc area being that which was to be later employed in extinction training. Responses to these areas were, therefore, at a higher practice level than those to the other area which was introduced on the first acquisition day. If area is to have a demonstrable effect upon latency, the situation should not involve disproportionate amounts of training. Moreover, plots of the acquisition latencies for each of the eight experimental groups indicate that by the end of the first acquisition day the latencies to both the extinction and the test areas were approaching the asymptote, suggesting that any latency differences which might be associated with area differences were being removed by the training procedure.

This interpretation of the reversal in the form of the gradients assumes that the inversion between Groups 79–79 and 79–50 is indeed a genuine phenomenon. The statistical analysis of the data does not support such an assumption, but because a comparable reversal has been obtained in an experiment employing similar *Ss* and procedures (**9**) there is some reason to accept, with caution, the hypothesis that the results are not due merely to sampling errors.

The general form of the gradients suggests that the hypothesis of falling gradients for generalized extinction effects in situations involving rather complex behavior sequences is tenable. However, the extreme variability of *Ss*' performances and the possible (unknown) distortion introduced by stimulus factors suggest caution in hypothesizing the "true" form of the primary stimulus-generalization gradients.

Considering the differences in procedure, the degree of correspondence between the gradients as determined by Thompson's procedure and the one used in the present study merits some comment. Thompson (**23**) tested each animal on each of four disc areas, whereas the present study used different groups of animals for each area. In addition, the response measures utilized were

in certain respects different: The latency recorded by Thompson was from the time the starting door was raised until the animal crossed a point 10½ in. from the starting compartment and 3¾ in. from the stimulus discs; while in the study here reported, the latency recorded was from the raising of the center door to the actual response to the stimulus. These measures may indeed be as different as were the "starting" and "running" times compared in Raben's (18) study of behavior on an elevated runway.

Despite these differences, there is a certain degree of similarity in form of the two sets of gradients in that they are both falling and both tend to have an intersection near their midpoints. The major differences appear to be the more rapid fall of the Thompson gradients and the appearance of the inversion in the gradient of response to stimuli of smaller area in the present study. Both the differences in the response measures and the testing procedures appear to be capable of producing these discrepancies.

## Summary

1. Albino rats were trained to obtain food pellets by opening a small door set in the center of a disc at the end of a reversible runway. Each *S* was given training on two discs (as, for example, 79 sq. cm. and 50 sq. cm. in area). Eight groups of 11 *S*s each were employed, each group being trained on one of the following area combinations: 79–20, 79–32, 79–50, 79–79; 20–79, 20–50, 20–32, 20–20.

2. Each *S* was given acquisition training consisting of 30 trials per day for three consecutive days, the trials being evenly distributed between the two areas utilized for his group. Extinction training was begun on Day 4, using only the first area of the training pair. Extinction was to the criterion of no response on 4 out of 5 consecutive trials. The *S* was required to again attain this criterion the following day, at which time he was immediately tested for generalization of extinction by presenting the second disc of the training pair.

3. Two generalization gradients were plotted in terms of the median latencies to the first response made following extinction; one gradient for *S*s extinguished on the largest disc area, and one for *S*s extinguished on the disc of smallest area. In general, the functions show the expected falling, negatively accelerated form.

4. The only departure from regularity and symmetry in the plots is a reversal in the expected trend between *S*s extinguished on the 79-sq. cm. disc and tested on (*a*) the 50-sq. cm. disc, and (*b*) the 79-sq. cm. disc. Whether this is a chance deviation or an indication of the greater stimulating value of larger stimulus area is not certain.

(Received April 15, 1952)

## References

1. Anrep, G. V. The irradiation of conditioned reflexes. *Proc. royal Soc. London*, 1923, **94B**, 404–425.
2. Bass, M. J., & Hull, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, **17**, 47–65.
3. Burros, R. H. Three rational methods for the reduction of skewness. *Psychol. Bull.*, 1951, **48**, 505–511.
4. Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
5. Ellson, D. G. Quantitative studies in the interaction of simple habits. I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.
6. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 97–105.
7. Grice, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. *J. exp. Psychol.*, 1948, **38**, 633–642.
8. Grice, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, **42**, 365–373.
9. Grice, G. R., & Saltz, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.
10. Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with

varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.

11. Hovland, C. I. The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.
12. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
13. Hull, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120–134.
14. Hull, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.
15. Hull, C. L. Simple qualitative discrimination learning. *Psychol. Rev.*, 1950, **57**, 303–313.
16. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
17. Pavlov, I. P. *Conditioned reflexes*. (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.
18. Raben, M. W. The white rat's discrimination of differences in intensity of illumination measured by a running response. *J. comp. physiol. Psychol.*, 1949, **42**, 254–272.
19. Schlosberg, H., & Stanley, W. C. A simple test of the normality of 22 distributions of electrical skin conductance. *Science*, 1952.
20. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
21. Spence, K. W. Failure of transposition in size-discrimination of chimpanzees. *Amer. J. Psychol.*, 1941, **54**, 223–229.
22. Spence, K. W. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, **31**, 257–271.
23. Thompson, M. E. The generalization of inhibition to stimuli varying in the visual size dimension. *Amer. Psychologist*, 1950, **5**, 253. (Abstract)
24. Wilcoxon, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.
25. Woodworth, R. S., & Schlosberg, H. *Experimental psychology*. (2nd Ed.) New York: Holt, 1952.
26. Youtz, R. E. P. The weakening of one Thorndikian response following the extinction of another. *J. exp. Psychol.*, 1939, **24**, 294–304.

# REMINISCENCE AS A FUNCTION OF THE AMOUNT OF CHANGE IN THE INTERTRIAL INTERVAL[1,2]

WARREN H. TEICHNER

*Climatic Research Laboratory, Lawrence, Mass.*

AND ELAINE HOLDER

*University of Missouri*

Several behavioral phenomena related to the length of the intertrial interval can be deduced from Hull's theoretical framework by appealing to the amount of reactive inhibition ($I_R$) present as a function of the amount of work performed or the length of the intertrial interval during which $I_R$ assumedly dissipates. For example, on the basis of differential dissipation of $I_R$, Hull's theory predicts that as the intertrial interval during conditioning is increased, conditioning becomes more rapid, but that when the interval is increased during extinction, extinction becomes slower. The first part of this prediction has firm support in the literature (e.g., **6, 8, 9**) but attempts to demonstrate more rapid extinction with massed extinction trials have resulted in conflicting data. Some studies (**7**) show less rapid and some (**4, 5, 6**) show more rapid extinction with massed trials. Still other studies (**3, 7**) indicate no difference at all in the rate of extinction under massed or distributed trials.

Reynolds (**6**) was the first to suggest the influence of the intertrial interval on the stimulus as a factor which might account for the conflicting results. In conditioning the eyelid response, Reynolds found that massed extinction trials (10–20 sec., averaging 15 sec. between trials) led to more rapid extinction than distributed extinction trials (60–120 sec. intervals, averaging 90 sec.) when the prior conditioning trials had been distributed, but that when the conditioning trials had been massed, there was no difference in extinction rates under the two conditions. To explain this Reynolds proposed a specific theory of the generalization of stimulus traces, the amount of generalization depending in part on the length of the intertrial interval.

In an instrumental conditioning situation, Teichner (**9**) found that the rate of extinction was definitely related to the degree of similarity between the intertrial interval employed during extinction and that used during conditioning. Thus, when the intertrial interval used during extinction was different from that used during conditioning, rapid extinction was favored, but when the same interval was used in both procedures, extinction tended to be slower.

Teichner accounted for the greater extinction rate with massed extinction trials in terms of the greater dissipation of reactive inhibition ($I_R$) during the longer intervals. To account for the effects of the similarity of the intertrial interval which produced results in conflict with the $I_R$ postulate, he followed Reynolds in appealing to the principle of stimulus generalization. Thus, the stimulus complex present at the beginning of any trial was said to depend in part on proprioceptive stimulus traces persevering from the pre-

[1] The data for this report were collected in the Psychological Laboratory of the New Mexico College of Agriculture and Mechanical Arts.

[2] Presented at the 1952 meeting of the Midwestern Psychological Association.

vious trial. To the extent that the intertrial interval is changed, the stimulus trace, and therefore the stimulus complex, should be changed, and by the principle of stimulus generalization, response strength should be less.

The same hypothesis has also been advanced by Montgomery (3) to account for his failure to find reminiscence when he increased the intertrial interval during conditioning following relatively massed practice. He, too, hypothesized that changes in the proprioceptive traces resulting from increasing the interval would effect a decrease in effective reaction potential ($_s\bar{E}_R$) sufficient to counteract the amount of $I_R$ which would dissipate during the interval. Although the Reynolds and Teichner studies provide support for the notion that both stimulus generalization and $I_R$ are influenced by the intertrial interval, neither of these studies was concerned with reminiscence. Montgomery's results, on the other hand, although deducible from this hypothesis, are not clear-cut in providing experimental support for it, since his data provide no evidence of reminiscence. In the absence of a demonstration of reminiscence varying in accordance with the operation of both principles, it is difficult to have high confidence in their generality. The experiment herein reported was designed for this purpose, i.e., to test the hypothesis that the effect of an interpolated intertrial interval on reminiscence is such that it can be deduced from (*a*) an increase in $_s\bar{E}_R$ due to a decrease in $I_R$ during the interval and (*b*) a decrease in $_s\bar{E}_R$ due to a change in the stimulus trace during the interval.

## Method

*Subjects.*—The *S*s were 72 naive albino and hooded rats from the colony maintained by the Department of Psychology, New Mexico State College. Sexes and strains were distributed equally among the six experimental groups. All *S*s were between 90 and 110 days old.

*Apparatus.*—The apparatus consisted of a flat black, straight alley with a starting box, a runway 56 in. long, and a goal chamber. The starting box and the goal chamber were separated from the runway by guillotine doors. Measurements of running time were taken from a point 4 in. beyond the starting-box door to a point 48 in. away. This latter point was 8 in. before the goal box. Measurements were recorded with the use of two photocell units and an electric clock measuring in 1/100 sec.

*Procedure.*—All *S*s were gentled and put on a 22-hr. feeding schedule for 7 days prior to the experiment. On the day prior to the experimental day, *S*s were put in the apparatus without reward, three at a time, and allowed to run freely for 15 min. with both doors raised. On the experimental day, *S*s were run following 22 hr. of food deprivation. With the restrictions of strain and sex noted above, *S*s were assigned to experimental groups by a table of random numbers prior to the experimental day.

On the experimental day each *S* was given 15 successive trials in the apparatus with "no" time between trials. This procedure consisted of allowing each *S* to run to a food reward of approximately .10 gm. of Purina Laboratory Chow and returning it by hand to the starting box as soon as the food was consumed. The actual time between trials depended, of course, on how quickly *S* ran on a particular trial and how quickly the food was consumed. In order to keep the time between trials at a minimum, a condition favoring the production of reminiscence, *S*s taking more than 15 sec. to pick up and eat the food in the goal box were discarded. In all, nine *S*s were discarded for this reason, leaving the total $n$ available for analysis at 63. Although this procedure is being referred to as "no time between trials," the actual time elapsed between the completion of one trial and the initiation of the next is estimated at 3 sec. on the average.

Following Trial 15 each *S* was placed in a small individual cage for a period of either 5, 15, 30, 60, or 180 sec. At the end of this period, *S* was returned to the starting box for Trial 16. In addition to the groups listed, one more group was run which was given 16 consecutive trials with "no" time between trials. The actual times between trials for all groups were approximately 3 sec. longer than indicated.

## Results

A simple analysis of variance performed on the reciprocals of the run-

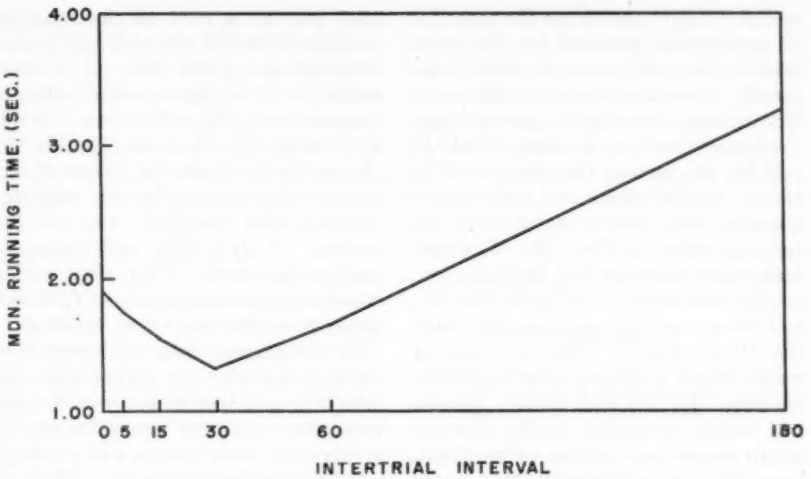


FIG. 1. Median running times on Trial 16 as a function of the time between Trials 15 and 16

ning times obtained from each group on Trial 15 yielded an $F$ ratio of less than 1.00, indicating that the discrepancies observed between groups could have easily arisen from simple random sampling. In addition, a plot of the median running times against groups showed no consistent trend, the form of the curve being wavelike, if anything.

The same kind of analysis of the data obtained from Trial 16 yielded an $F$ of 2.40 which with 5 and 58 $df$ is significant at the 5% level of confidence. In Fig. 1 is shown the median running times as a function of the time interval between Trials 15 and 16. Median running times decreased successively for each time interval up to 30 sec., the form of the function up to this point being apparently negatively accelerated. Following 30 sec. the curve rises fairly rapidly up to 180 sec., the longest interval used.

An important fact shown in Fig. 1 is that the median running time for the 60-sec. group is less than that for the 0- and 5-sec. groups, but greater than that for the 30-sec. group. In other words, although the curve is decreasing from 0 to 30 sec., a gradual reversal in trend is also apparent. This reversal is emphasized by the long running time of the 180-sec. group. If the effect of the variation of the intertrial interval were a simple one, i.e., comparable to the variation of a single variable, a simple monotonic function might be expected. Since this is not what was obtained, it is more reasonable to suppose that the curve depends on the joint variation of two or more variables each of which is a function of the intertrial interval. In any case, it is clear that an increase in the intertrial interval produced first a decrease in running times and then an increase.

## DISCUSSION

As formulated by Hull, $I_R$ accumulates with the occurrence of each response and dissipates as a function of the time between responses (2). In the present experiment the amount of such inhibition accumulated by the

end of Trial 15 should be the same for all groups and maximal for the effort involved in the response since completely massed trials were employed. Accordingly, it would be expected that the longer the time between Trials 15 and 16, the greater the amount of $I_R$ which would dissipate, and consequently, the faster should be the running times on Trial 16. The running times actually did decrease, i.e., exhibit reminiscence, in the 5-, 15-, 30-, and 60-sec. groups, as compared with the 0-sec. group. The reversal of trend which becomes most apparent between the 30- and 60-sec. groups and which continues to the 180-sec. group is contrary to this expectation, however. It is apparent that some factor was operating to depress the amount of reminiscence. In order to account for this, a second principle may be invoked, one which can operate in competition with $I_R$ in determining ${}_s\bar{E}_R$. Since, within the present experiment, it is not reasonable to assume differences in drive strength between groups, the only alternative within Hull's framework is the construct of habit strength (${}_sH_R$).

${}_sH_R$, which is defined in terms of the number of reinforced responses, can decrease only under conditions in which there has been a change in the stimulus trace with which it is related. The manner in which ${}_sH_R$ decreases under these conditions is called the gradient of stimulus generalization. Since the stimulus trace on which this gradient depends is a function of the time since the termination of the stimulus, other factors constant, a change in the time since the termination of the stimulus should produce a change in the stimulus trace and consequently a decrease in effective habit strength (${}_s\bar{H}_R$). If the intertrial interval is not too long, the stimulus traces persevering from the previous trial will be a part of the stimulus complex to which the response is conditioned on a given trial, or, in other words, it is to this specific stimulus complex that ${}_sH_R$ will accrue. It follows from this that the greater the change in the intertrial interval, the greater the change in the stimulus complex and, therefore, the less the amount of ${}_s\bar{H}_R$, ${}_s\bar{E}_R$, and measured response strength. This, then, would counteract the dissipation of $I_R$ which depends on the same time interval.

In the present data, the reversal of trend found with the longer intertrial intervals is taken as support for the view that stimulus generalization is involved in every instance of a change in the intertrial interval. If it is assumed that $I_R$ decreases with time more rapidly than ${}_s\bar{H}_R$, then ${}_s\bar{E}_R$ would be more greatly affected by the decrease in $I_R$ and less by the loss in ${}_s\bar{H}_R$ where short time intervals are concerned. For longer intervals, involving greater changes in the intertrial interval, there would be relatively little further increase in ${}_s\bar{E}_R$ due to further decreases in $I_R$, but in this case ${}_s\bar{E}_R$ would be more greatly affected by the greater loss in ${}_s\bar{H}_R$ concomitant with the change to the longer intervals.

On the above basis it would be expected that following closely massed trials, response strength would increase (reminiscence) with increases in the intertrial interval up to some point at which the increase would begin to lessen, and finally the strength of response would actually decrease. This, of course, is exactly what happened in the present experiment. It would also follow that if the trial spacing during acquisition were sufficiently long (allowing practically no accumulation of $I_R$) a change in the intertrial interval should result in a decrease in response strength. Some

data supporting this prediction are available in the area of human motor learning (1). The prediction can also be made that an increase in the interval following relatively, but not closely, massed trials would result in either no change or a drop in the strength of response, depending on the relative contributions of $I_R$ and ${}_s\bar{H}_R$ to ${}_s\bar{E}_R$. Montgomery's (3) failure to obtain reminiscence is now easily understood since his most highly massed condition involved a 20-sec. intertrial interval and this was changed to a 24-hr. interval. As Montgomery points out, the effects of generalization were presumably sufficient to overcome the dissipation of $I_R$.

On the basis of the above considerations, it appears reasonable to extend the generality of Hull's theoretical principles as they have been used by recent writers to account for the effects of changes in the intertrial interval. This is deemed reasonable since the present experimental investigation found the amount of reminiscence as a function of the amount of change in the intertrial interval to be consistent with the postulated operation of both the principle of stimulus generalization and of reactive inhibition.

## Summary

Sixty-three rats were run in a straight alley for 15 massed trials. Intertrial intervals of 0, 5, 15, 30, 60, and 180 sec. were then interpolated between Trials 15 and 16. The amount of reminiscence obtained on Trial 16 was found to be a complex function of the interpolated interval, response strength first increasing as the interval became longer and then decreasing with still longer intervals. These results were found consistent with Hull's principles of reactive inhibition and stimulus generalization, especially as they have been employed in recent reports to account for the effects of changes in the intertrial interval.

(Received April 18, 1952)

## References

1. Ammons, R. B. Acquisition of motor skills: III. Effects of initially distributed practice on rotary pursuit performance. *J. exp. Psychol.*, 1950, **40**, 777–787.
2. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
3. Montgomery, K. C. An experimental investigation of reactive inhibition and conditioned inhibition. *J. exp. Psychol.*, 1951, **41**, 39–51.
4. Razran, G. H. S. The nature of the extinctive process. *Psychol. Rev.*, 1939, **46**, 264–297.
5. Reyna, L. J. Experimental extinction as a function of the interval between extinction trials. Unpublished Ph.D. thesis, State Univer. Iowa, 1946.
6. Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
7. Sheffield, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.
8. Spence, K. W., & Norris, E. B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.
9. Teichner, W. H. Experimental extinction as a function of the intertrial intervals during conditioning and extinction. *J. exp. Psychol.*, 1952, **44**, 170–178.

# EFFECTS OF ELECTROSHOCK CONVULSIONS ON LATENT LEARNING

A. W. BENDIG AND R. A. PATTON[1]

*University of Pittsburgh, Western Psychiatric Institute and Clinic*

The accumulating evidence on the effects of electroshock convulsions (ESC) on the learning and retention of nonhominid organisms suggests that a most important dimension is the complexity of the task to be learned or retained. The learning and retention of relatively simple maze habits seem little affected by ESC (**5**), but more complex types of maze problems have rather consistently shown decrements in learning and retention after ESC. Habit-reversal studies have shown that the effect of ESC is positively related to the difficulty of the task (**2**).

One aspect of learning behavior in the rat that appears to operate at a more complex level than tasks previously used in ESC research is the behavior called "latent learning." The recent reviews of the controversial concept of latent learning testify to an interest in, and argument about, the mechanisms operative in this phenomenon (**4, 6, 7**). The interest of the present authors was in developing a situation in which latent learning would occur and to measure the effects of ESC on this learning.

In a previous article (**1**) it was demonstrated that a water-maze problem could be developed to measure latent learning in the rat. The maze used was a relatively simple water maze with one choice point and two metal landing platforms. The experimental procedure used (which is briefly recapitulated below) can be found in the previous report (**1**).

## METHOD

*Subjects.*—The *S*s used were 45 male white rats of the Albino Farms strain. They were randomly divided into an experimental group of 22 and a control group of 23. The rats were 116 to 128 days of age at the beginning of this experiment and were experimentally naive.

*Apparatus.*—The maze described in a previous report (**1**) was used. Briefly, it is a single choice-point **T** maze filled with water to a depth of 8 in. The two landing platforms were of wire mesh and retrace doors were located at the choice point. The apparatus used for administering the electroconvulsive shock was the Pittsburgh Electroshock Unit (**2**).

*Procedure.*—The *S*s of both groups were given 20 adaptation trials in a water-filled straightaway distributed as four trials for each of five days. On these trials *S* was required to swim a 4-ft. straightaway to a wire-mesh landing platform at the opposite end. These trials were given to accustom *S*s to the swimming task required of them in the experimental phase of the study.

After completion of the adaptation trials both groups were given four trials in the maze per day for four days. For this four-day training period food and water were available at all times in the home cages of all *S*s. During this period, half of the *S*s in each group found food pellets piled on the right-hand landing platform and none on the left; the remaining *S*s consistently found food always on the left platform and never on the right. The first and third trials each day constituted a free-choice situation in that *S* was free to choose a right or left turn at the choice point. The second and fourth trials were forced choices, *S* being forced by a closed retrace door at the choice point to turn in the opposite direction from his free choice on the immediately preceding trial. This procedure was followed to insure that each *S*'s total experience with the food-filled and empty landing platforms would be similar in both frequency and recency.

One major difference distinguished the experimental from the control group during this training period: each *S* of the experimental group received one electroshock convulsion each of the

[1] This experiment was performed in the Laboratory of Neurophysiology, Western Psychiatric Institute and Clinic, and was supported by a research grant from the U. S. Public Health Service.

four days, whereas the control group *Ss* never were convulsed. The ESC was administered 1 hr. after the completion of the fourth training trial each day and the shock was given at a level of 20 ma. for .1 sec.

After the fourth-day trials, all *Ss* were replaced in their cages from which all food had been removed, but in which water was still available, and 23 hr. later the *Ss* of both experimental (ESC) and control groups were given four free-choice trials in the maze. On these trials food was found by all *Ss* on both of the water-maze landing platforms. On each trial *S* was scored as making or not making a food-side response, i.e., turning or not turning toward the landing platform where he had always found food present during the training trials.

## Results

The number of *Ss* making food-side and non–food-side responses on each trial on the fifth day can be found for both ESC and control groups in Table 1.

The frequencies of animals making each response in the two groups were tested for significance by setting up a fourfold contingency table for each trial and computing chi square from the distributions of frequencies. A declining gradient of significance can be seen over the four trials with the control group making significantly (.02 level) more food-side responses than the ESC group on the first trial. However, this superiority of the control group was eliminated by the third trial, and the fourth trial showed a very small chi-square value.

Since an increasing preference for the food side during the four training days had been noted in a previous study (1), the percentages of food-side responses of each group on the two free-choice trials of each of the four training days and the first two trials of the fifth day were graphed and can be found as Fig. 1. An increasing preference for the food side can be noted in the control group, but the responses of the ESC group tend to fluctuate randomly between food and non-food responses.

TABLE 1

Number of *Ss* Making Food-Side and Non-Food-Side Responses on the Fifth Day

| Trial | ESC Group | | Control Group | | $\chi^2$ | $p$ |
|---|---|---|---|---|---|---|
| | Food | Non-food | Food | Non-food | | |
| 1 | 9 | 13 | 18 | 5 | 6.536 | .02–.01 |
| 2 | 9 | 13 | 17 | 6 | 5.021 | .05–.02 |
| 3 | 13 | 9 | 17 | 6 | 1.112 | >.05 |
| 4 | 11 | 11 | 14 | 9 | 0.538 | >.05 |

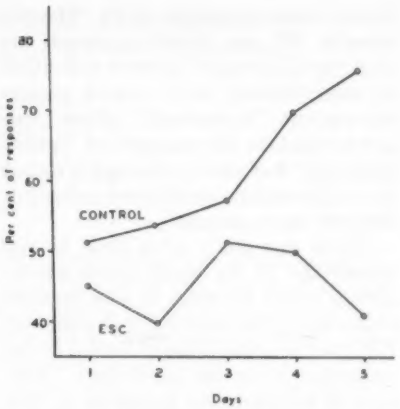


Fig. 1. Percentages of food-side responses on the free-choice trials on the four training days and on the fifth (test) day

## Discussion

From the evidence presented above it appears legitimate to conclude that electroconvulsive shock had a seriously detrimental effect on the operationally defined "latent learning" of this study. Despite the growing preference of the control group for food-side responses during the four training days, it is not difficult to defend the assumption that the learning shown by *Ss* in the control group was "true" latent learning. Previous studies have reported a similar improvement in maze performance in animals during the experimentally defined "no-reward" phase of

latent-learning studies (1, 3). Thistlethwaite (7, pp. 65–66) pointed out that the occurrence of error reduction in experimental and control groups during the "no-reward" phase does not invalidate the concept of "latent learning," but such a concept is tested by a differential *rate* of error reduction between such groups.

While this study adds little to our knowledge of the mechanisms operative in latent learning, it does provide evidence of the detrimental effect of ESC on the acquisition of habits under irrelevant incentive conditions. One area of uncontrolled variables in this study, as in most ESC research, is the effect of ESC on motivation. The hunger drive present in the control animals during training may not have been of equal strength in the ESC group during the same period. Until experimental evidence is presented on this problem, our results, and previously reported results in the field of ESC and behavior, must be somewhat tentative. However, we can conclude that a food-side preference acquired by rats during a "non-reward" training period does not appear in animals undergoing electroshock convulsions. However, this food-side preference of the control group rapidly disappeared when given primary reward for both directional responses at the choice point and the animals of this group rapidly acquired the random-choice behavior of the ESC group on the later trials of the hunger-motivated fifth day.

## Summary

Forty-five white rats were given 20 adaptation trials in a water-filled straightaway. They were then given 16 trials at the rate of four per day in a single choice-point water maze while on an ad libitum food and water regimen. Food was always present on one of the two landing platforms in the maze. The first and third trials each day were free-choice trials, the second and fourth trials forced choices to equate food-side and non–food-side experience. One group of 22 *S*s was given one electroconvulsive shock 1 hr. after the last trial on each of the four days; a control group of 23 *S*s received no convulsions. On the fifth day both groups operated under 23-hr. food deprivation and were given four free-choice trials in the maze with food reward present on both landing platforms.

The control group made significantly more food-side responses on the first two trials of the fifth day, but approximated the random-choice behavior of the ESC group on the last two trials. This initial food-side preference of the control group can be attributed to an increasing food-side preference displayed by the control group during the non–hunger-motivated training period. The ESC group did not show a food-side preference during the training or testing periods.

(Received April 19, 1952)

## References

1. Bendig, A. W. Latent learning in a water maze. *J. exp. Psychol.*, 1952, **43**, 134–137.
2. Braun, H. W., & Patton, R. A. Habit reversal after electroshock convulsions as a function of the difficulty of the task. *J. comp. physiol. Psychol.*, 1950, **43**, 252–263.
3. Meehl, P. E., & MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, **41**, 372–396.
4. Postman, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489–563.
5. Russell, R. W. Effects of electroshock convulsions on learning and retention as functions of the the difficulty of the task. *J. comp. physiol. Psychol.*, 1949, **42**, 137–142.
6. Thistlethwaite, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.
7. Thistlethwaite, D. Reply to Kendler and Maltzman. *Psychol. Bull.*, 1952, **49**, 61–70.

# THE RELATIONSHIP BETWEEN TWO KINDS OF INHIBITION AND THE AMOUNT OF PRACTICE

GREGORY A. KIMBLE

*Duke University*

AND ROBERT B. SHATEL[1]

*Brown University*

In a previous paper an attempt was made by Kimble (4) to use Hull's concepts (2), reactive inhibition ($I_R$) and conditioned inhibition ($_sI_R$), for explanation in the field of motor learning. The theoretical argument developed in that paper demonstrated, among other things, that Hull's theory, applied to motor learning, leads to testable deductions concerning the course of development of $I_R$ and $_sI_R$. It was predicted that the curve relating $I_R$ to the amount of practice would rise quickly to a maximum and maintain this maximal level as long as motivation remained constant. It was predicted, further, that the analogous function for $_sI_R$ would be negatively accelerated and increasing.

The same paper described a procedure for getting measures of the two key concepts in the theory and tested these predictions in an experiment where *S*s printed the alphabet upside down and backwards. In general, the experimental results supported the predictions derived from the theory, but only in general. The measure of $_sI_R$ showed a tendency toward the predicted negatively accelerated form, but failed to reach an asymptote. The measure of $I_R$ attained the predicted maximum after several minutes of practice, but showed the tendency, also found by Ammons (1) and Irion (3), to decrease later on in learning. Thus, both of these results are encouraging; but both also indicate the advisability of checking them in a study covering a larger amount of practice. The present experiment was performed to make such a check. It uses a modification of the same method and a considerably greater number of learning trials on the assumption that a more nearly complete picture of the two relationships may be obtained under these conditions.

[1] This article is an abridgement of the junior author's Master's thesis. We wish to thank Prof. C. I. Hovland for a critical reading of the manuscript.

## Method

*Apparatus.*—The apparatus used in this experiment was a pursuit rotor, which ran at a rate of 78 rpm. The turntable was 16 in. in diameter and carried a silver target the size of a dime set flush with its surface, 5 in. from the center. The faster than usual rotation speed and the greater distance per revolution through which the target traveled were introduced to make the task more difficult and to insure learning over at least the majority of the learning trials. In other respects the apparatus was like that used by Kimble (5).

*Subjects and procedure.*—Sixteen male and four female college students, naive to rotary-pursuit learning, were *S*s in the experiment. Two women and eight men were assigned at random to a massed- and a spaced-practice group.

Both groups of *S*s practiced on the rotor for 15 trials a day, 5 days a week, for two successive weeks, and a total of 150 trials. For the massed group, each day's practice session consisted of 15 50-sec. trials separated by 10-sec. pauses during which the timer was read and reset. In the case of the spaced group, each practice session involved 15 50-sec. trials separated by 70-sec. pauses. The *S* began tracing the target during the last 5 sec. of the intertrial interval; so the intertrial rests for the massed- and spaced-practice groups were actually about 5 sec. and 65 sec., respectively. Thus the two groups worked on identical schedules except for the difference in the duration of the intertrial rest period. The

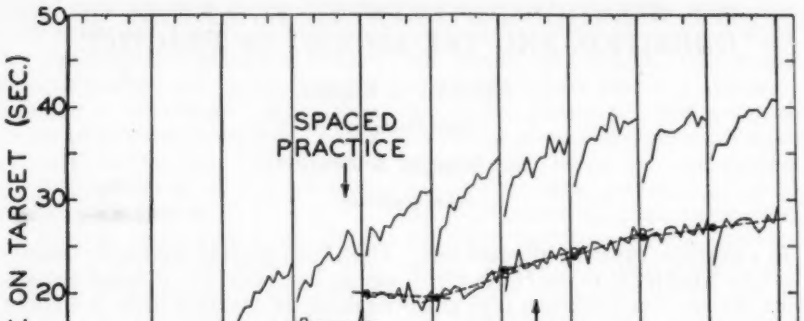


FIG. 1. Learning curves for the two experimental groups. The dashed lines were fitted by the method of averages to the relatively decremental portion of the massed-practice functions. The closed and open circles are, respectively, estimates of what performance might have been on the first trial had there been no rest, and had there been no need to warm up.

*Ss* were given their score at the end of each trial and were aware of the number of trials per session as well as the total number of practice sessions.

## RESULTS

The basic data obtained in our experiment are shown in Fig. 1 which presents learning curves for the two groups. The values plotted are mean scores for each of the 150 50-sec. trials. The breaks in the curves correspond to the intersession rests of about 24 hr. One notes immediately the typical superiority of performance in the spaced-practice condition.

The dashed straight lines through the curves for the massed-practice group were fitted by the method of averages to the *relatively* decremental portions of these functions for each session. The identification of the relatively decremental portion of the curve, of course, involves a judgment. But the problems thus presented were not serious. In almost every case the fit is to the last 14 points for each session. The exceptions are in the first, fifth, and tenth sessions where the fit is for the last 13 points. By extrapolating these curves backward to the first postrest trial we obtain, as Ammons (1) has pointed out, an estimate of what performance might have been on the first trial of each session had there been no need to "warm up." And, by extrapolating the function *forward* to the first trial of the *following* session, we arrive at an estimate of what performance would have been on that trial had there been no rest. These estimated values are used in the derivation of Fig. 2 and 3, which present the functions this experiment was designed to obtain.

The index of *reminiscence* plotted in Fig. 2 is the algebraic difference be-

tween the actual score of the massed-practice group on the first postrest trial of each session and the estimated score (open circle in Fig. 1) for the same group on that trial. The striking feature of this function is that, after 30 trials, the amount of reminiscence begins to decrease and is actually negative for the sessions late in the experimental sequence. A series of $t$ tests for related measures (testing the hypothesis that the mean gain or loss from the last trial on one session to the first trial on the next session is zero) was run, in order to examine the gross dependability of the reminiscence function in Fig. 2. The $t$ values, thus computed, yield acceptable levels of significance ($p = .05$ or better for two tails) only for the gains after Trials 15 and 30 and for the losses after Trials 120 and 135. This is taken as verification of the *general* relationship presented in Fig. 2. That is, under these experimental conditions, rest produces reminiscence early in learning and later on produces a performance loss.

A re-examination of Fig. 1 will reveal that these performance losses occur because of a warm-up decrement which depresses performance for one or two trials in each session. The measure of $I_R$ plotted in Fig. 2 supports this interpretation, since it is based upon an estimate of performance *corrected for warm-up decrement.* The specific measure of $I_R$ presented in Fig. 2 is the difference between the estimate of performance on Trial 1 of each session had there been no rest (open circle in Fig. 1) and the estimate of performance on the same trial had there been no need to warm up (closed circle). This index of $I_R$ is at a maximum after 15 and 30 trials. Then it decreases rapidly, reaching zero after 75 trials and remaining at approximately this level for the rest of practice. This tendency for the $I_R$ function to stabilize at zero supports the suggestion that negative values of reminiscence occur entirely because of the warm-up decrement.

The purpose of running the spaced-practice group was to provide a means of tracing the development of ${}_sI_R$. The total difference between the performance levels of massed- and spaced-practice Ss is thought of as representing the sum of a temporary decrement ($I_R$) and a permanent one (${}_sI_R$). The latter component of this difference is what remains of the total after reminiscence and a correction for warm-up decrement have been removed by subtraction. Figure 3 shows the development of ${}_sI_R$. The measure plotted there is the difference between the score estimated for the massed-practice group on the first trial of each session had there been no need to warm up and the actual score of the

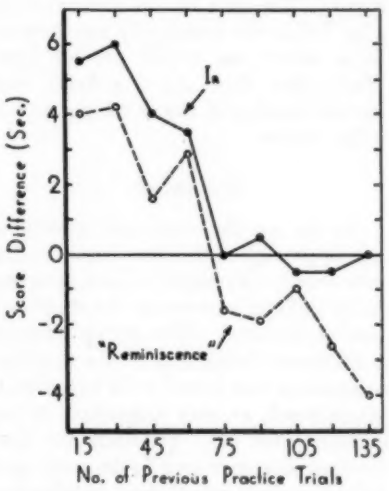


FIG. 2. Amount of $I_R$ and amount of reminiscence as a function of the number of previous practice trials. See text for method of determining plotted values.

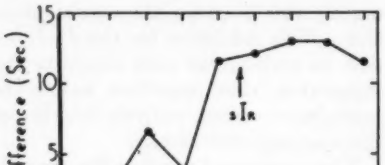


FIG. 3. Amount of $_sI_R$ as a function of the number of previous practice trials

spaced-practice group on the last trial of the preceding session. The use of the last score on the preceding trial for the spaced group seemed to provide a better measure of performance at any given stage of practice than the first score on the next session because of the serious distortion of these latter points resulting from warm-up decrement. The plot of $_sI_R$ presented in Fig. 3 has the negatively accelerated form which one would predict from Hull's idea that $_sI_R$ is a *habit*, and should develop in the same manner as other habits.

## DISCUSSION

So far as the empirical questions which prompted this investigation are concerned, they seem to have received fairly definite answers in the results of this experiment. The curve relating conditioned inhibition to the amount of practice was found to be negatively accelerated, as was suggested in an earlier report (**4**). Similarly the fact that reminiscence and $I_R$ decrease late in learning was established, confirming and extending results previously reported (**1, 3, 4**). In general these findings support Kimble's (**4**) theory of motor learning, although the finding that $I_R$ reduces to zero after considerable training suggests a reconsideration of one point in the theory.

Such decreases in $I_R$, according to Kimble, are attributable to a decrease in the threshold of $I_R$ an $S$ will tolerate before resting occurs. And the one mechanism described for reducing this threshold is a decrease in motivation. Implicit in the theory, however, is an alternative mechanism which seems worthy of consideration. Values of $I_R$ above the threshold for resting have always been treated by Kimble as unconditioned *stimuli* which elicit a resting response. The weaker values of $I_R$ are presumably also stimuli which are not strong enough to possess drive properties, *but may have cue value.* Furthermore, since these weaker values of $I_R$ always precede the resting response and its reinforcement, the conclusion that resting behavior will get conditioned to the weaker values of $I_R$ follows from conditioning theory. This, then, is the second method by which the threshold for resting may be reduced, since to say that resting responses are conditioned to weak values of $I_R$ is the same as saying that $S$ will tolerate less reactive inhibition before resting occurs. The chief advantage of this explanation is that it makes decrements in $I_R$ possible when motivation is constant. Whether one or both of these mechanisms is necessary defines a question for further research.

## SUMMARY

In an experiment designed to determine the shapes of the curves relating reactive inhibition and conditioned inhibition to the amount of practice, 20 *S*s were given ten daily practice sessions and a total of 150 50-sec. trials on the pursuit rotor. Half of

the *Ss* rested about 5 sec. between trials; the other half rested for 65 sec. between trials.

Using calculational procedures developed in a previous paper (**4**) it was found that reactive inhibition decreases to zero late in practice and that conditioned inhibition increases with practice in a negatively accelerated fashion. These results confirm and extend those previously reported. They are interpreted in terms of Kimble's theory of motor learning.

(Received April 22, 1952)

## References

1. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
2. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
3. Irion, A. L. Reminiscence in pursuit rotor learning as a function of length of rest and of amount of pre-rest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.
4. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
5. Kimble, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.

# TWENTY QUESTIONS: EFFICIENCY IN PROBLEM SOLVING AS A FUNCTION OF SIZE OF GROUP[1]

DONALD W. TAYLOR AND WILLIAM L. FAUST

*Stanford University*

"Twenty Questions," popular as a parlor game in earlier years and now popular as a program on both radio and television, involves a type of problem solving that is of considerable interest psychologically.[2] To start the game, the participants are told only whether the object they are to attempt to identify is animal, vegetable, or mineral. In searching for the object which is the solution to the problem, they ask a series of questions, each of which can be answered "Yes" or "No." To find the solution most economically, they must use a high order of conceptualization, gradually increasing the specificity of the concepts employed until they arrive at the particular object.

The game is of psychological interest first of all because it appears to involve a type of problem solving more similar to much problem solving in everyday life than that ordinarily studied in psychological experiments. The solution is obtained not by a series of rigorous well-defined steps. Rather one starts with a general, somewhat vague problem. Questions are asked and information obtained. Upon the basis of this information, new questions are formulated. This procedure continues until the problem is solved. This type of problem solving is also of interest because it seems more similar to much of the problem solving in scientific research than does that involved in problems susceptible of rigorous, deductive mathematical or logical solution.

The use of the game in psychological experiments is recommended by several other considerations: It is quite interesting to college undergraduates; motivation is easily sustained for a period of several days. A very large number of problems of this kind are available. The same problems can be used with children and with adults. The same problems are appropriate for use with individuals and with groups of varying size.

The present experiment, the first in a series planned using the game, was designed to answer three questions: (*a*) How rapidly is the skill involved in the game learned? (*b*) How does efficiency in solving this type of problem vary as a function of the size of the group participating? (*c*) Does improvement in individual performance occur more rapidly with individual practice or with practice as a member of a group?

The second of these three questions is perhaps the most interesting. For many kinds of work, it seems quite reasonable that if a particular job must be completed in a shorter time, the number of people in the group working on it should be increased. It

[1] This experiment was carried out under Project NR 192-018 supported by Contract N6 onr-25125 between the Office of Naval Research and Stanford University. The first author designed the experiment, supervised the analysis of the data, and prepared the present report. The second author conducted the experiment and carried out the analysis of the data. Work on the contract is under the direction of the first author.

[2] The idea of using "Twenty Questions" in experimental studies of problem solving is not new. As was discovered after the present study was partly completed, Lindley (**3**) suggested the use of the game for this purpose in an article published in 1897.

is not clear that increasing the size of a group engaged in solving a problem will necessarily reduce the time required for its solution. Indeed, it appears likely that in some cases it will actually increase the time required. Shaw (**4**) has presented data which indicate that the performance of groups of four is superior to that of individuals. However, further experimentation with larger samples, varying size groups, and different types of problems is needed to determine adequately the relation between group size and efficiency in problem solving.

## Procedure

A total of 105 students from the elementary course in psychology served as *Ss*. The *Ss* were assigned by chance to work in solving the problems either alone, in pairs, or as a member of a group of four. There were 15 individual *Ss*, 15 groups of two, and 15 groups of four. Each individual or group was given four problems a day for four successive days. On the fifth day, all *Ss* worked alone, each being given four problems.

From a longer list of objects originally constructed, 60 were selected for use as problem topics. Included were 20 animal, 20 vegetable, and 20 mineral objects. Excluded were objects which did not clearly fit in only one of the three categories; e.g., hammer was not included because, with a handle of wood and a head of metal, it would be classed as both vegetable and mineral. Also excluded were objects which could not be expected to be familiar to almost every college student. Examples of objects included are: newspaper, Bob Hope, scissors, camel, dime, rubber band.

With four problems a day for five days, a total of only 20 problems was needed for presentation to any particular *S* or group. However, to minimize the possibility that an *S* would have any knowledge of what problem object to expect, it was decided to use a total of 60 different objects. This precaution seemed desirable although the instructions to be given all *Ss* specifically requested that they not discuss the problems with other students. It should be added that no evidence was obtained during the course of the experiment to indicate that any *S* had previously heard mentioned a problem object he was to be given.

Since the nature of the learning curve was of interest, it was necessary to control the order of presentation of the problems in such a way that those given on any one day would be equal in difficulty to those given on any other day. In the absence of any measure of the difficulty of the individual problems, the following procedure was employed: The 20 animal objects were listed in chance order, as were the 20 vegetable and the 20 mineral objects. To obtain a group of four for use the first day, the first item was taken from each of the three lists together with the next item from one of the three chosen by chance. Similarly, to obtain four objects for use the second day, the next item was taken from each of the three lists; the fourth item was then obtained by taking the next in order on one of the two lists from which the extra item had not been taken the first day. This procedure was repeated to provide four problems for the third, fourth, and fifth days. A second and a third set of four problems for each of five days were obtained by continuing the same procedure. Next the three lists of 20 were individually reshuffled and the entire procedure repeated to obtain a fourth, fifth, and sixth set.

In the experiment, the first, seventh, and thirteenth individual, pair, or group of four *Ss* received the first set of problems. The second, eighth, and fourteenth received the second set, and so on. As a result of this procedure, the order and the frequency of appearance of the problems were the same for individual *Ss* as for groups of two or of four.

All *Ss* were told that both the number of questions and the time required to reach solution would be recorded, but it was emphasized that number of questions was the more important score. In presenting each problem, *E* stated simply whether the object sought was animal, vegetable, or mineral. Time was measured by means of a stopwatch. A special data sheet was used for groups of two and of four to record which *S* asked each question. To each question, *E* replied "Yes," "No," "Partly," "Sometimes," or "Not in the usual sense of the word." If the question could not be answered in one of these ways or was unclear, *S* was asked to restate it.

The instructions given to groups of two or of four made clear that they might talk freely to each other, reviewing answers to previous questions or suggesting possible questions to ask. It was emphasized that they were not to compete against each other, but were to cooperate as a group to get the answer; they were told that the efficiency of their group would be compared with that of other groups.

As the name of the game indicates, *Ss* are traditionally allowed 20 questions in which to obtain the solution. Pretesting showed, however, that with naive *Ss* this limit results in a rather large proportion of failures. Accordingly,

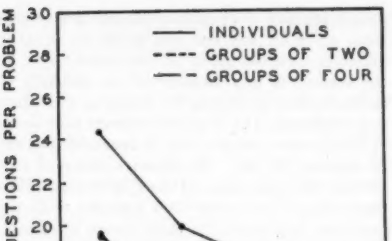


FIG. 1. Number of questions per problem as a function of days of practice and of size of group

to simplify the analysis of the data to be obtained, the number of questions permitted was increased to 30. Examination of the distributions of scores obtained suggests that, at least after the first day, the performance of individuals or groups of *S*s who do not reach solution in 30 questions is qualitatively different from that of those who do. The *E*'s impression is that in most cases of failure there was established an incorrect set which was unchanging even in the face of answers irreconcilable with it; it seemed that in such cases the *S*s might easily have asked 50 or 60 questions without solving the problem.

## RESULTS

*Rate of learning.*—The first question the experiment was designed to answer concerned the speed of learning of the skill involved. The data in Fig. 1 show that there is rapid improvement in the performance of both individuals and groups. By the fourth day the curves appear already to be flattening out. The score for an individual or single group for one day was the median of the number of questions required to solve each of the four problems on that day. The median was used instead of the mean because there were some failures. Each point plotted in Fig. 1 is the mean of these median scores on one day for 15 individuals, or for 15 groups of two or of four. In those few cases where an individual or group failed two or more problems on a single day, the median was obtained by treating the failures as though solution had been reached in 31 questions; the number of such cases was too small to affect the results appreciably; after the first day there were no such cases except among individual *S*s and even there they were rare.

The mean number of failures per problem on each day by individuals or groups is shown in Fig. 2. Thus, for example, on the first day the mean number of failures per problem among the 15 groups of four was .08; in other words, about one-twelfth of the problems were failed. The improvement in performance over four days in terms of number of failures per problem is consistent with that shown in Fig. 1 in terms of number of questions per problem solved.

Figure 3 shows the decrease over four days in the amount of time

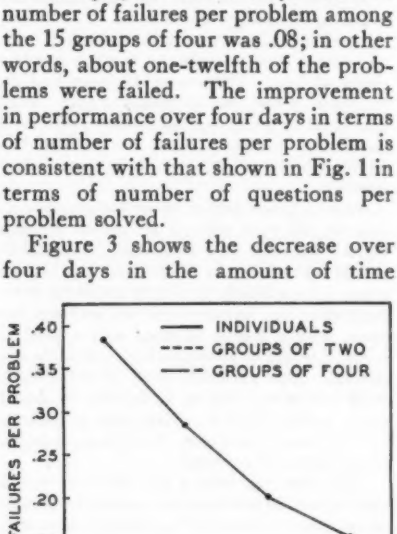


FIG. 2. Number of failures per problem as a function of days of practice and of size of group

required per problem. The time required, of course, is somewhat dependent on the number of questions asked, although not entirely so. The score for an individual or single group for one day was the median time required for solution of the four problems. In those few cases where there were two or more failures in one day, the median of the four times was taken simply as obtained; this procedure underestimates somewhat the median time that would have been required to solve all four problems, but as before the number of such cases was too small to affect the general results appreciably.

*Size of group.*—The second and major question with which the experiment was concerned involved the relation between efficiency in problem solving and size of group. As is evident in Fig. 1, there was no significant difference between groups of two and groups of four in terms of the number of questions required to reach solution. The performance of individuals working alone, however, was consistently inferior to that of either size group. The *t* technique was used to test the difference on each day between the mean score of the 15 individuals and the mean score of the 15 pairs of *Ss*, and also that of the 15 groups of four. The values of *t* obtained are presented in Table 1. With 28 *df*, a *t*

TABLE 1

VALUES OF *t* FOR DIFFERENCES BETWEEN MEAN SCORES: NUMBER OF QUESTIONS PER PROBLEM

| Day | Individuals versus Groups of Two | Individuals versus Groups of Four |
|---|---|---|
| 1 | 2.67 | 2.18 |
| 2 | 2.86 | 1.96 |
| 3 | 2.30 | 2.22 |
| 4 | 2.11 | 2.4 |
| All 4 | 2.64 | 2.62 |

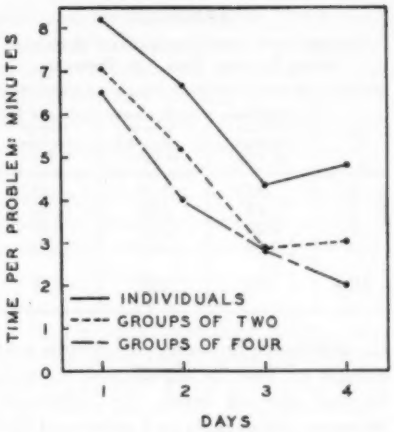


FIG. 3. Time per problem as a function of days of practice and of size of group

of 2.05 is required for significance at the .05 level and of 2.76 at the .01 level. All of the differences but one are significant at or beyond the .05 level.

A score for all four days was obtained for each individual or single group by taking the median number of questions required to solve the 16 problems. In terms of the means of these scores, the performance both of groups of two and of four is significantly better (.02 level) than that of individuals working alone (see Table 1).

That there were differences as a function of group size in terms of number of failures to reach solution is suggested by Fig. 2. Because of the fact that, as would be expected, the distributions of failure scores were not normal, *t* could not be used to test the significance of these differences. Instead a test described by Festinger (2) was employed. The mean number of failures per problem, all four days included, was for individuals, .26; for pairs, .10; for groups of four, .03. The values of *d* obtained indicate that

TABLE 2

VALUES OF $t$ FOR DIFFERENCES BETWEEN MEAN SCORES: TIME PER PROBLEM

| Day | Individuals versus Groups of Two | Individuals versus Groups of Four | Groups of Two versus Groups of Four |
|---|---|---|---|
| 1 | .85 | 1.14 | .12 |
| 2 | 1.01 | 2.36 | .93 |
| 3 | 2.20 | 2.22 | .06 |
| 4 | 2.15 | 3.49 | 1.90 |
| All 4 | 2.39 | 3.27 | 1.18 |

the difference between individuals and groups of four is significant at well beyond the .01 level; the difference between individuals and pairs and the difference between pairs and groups of four are both significant at about the .02 level.

Differences in mean time to solution among individuals, groups of two, and groups of four may be seen in Fig. 3. Fortunately, the distributions of the median times, of which the individual points plotted in Fig. 3 are the means, were such as to make the use of $t$ appropriate in testing the significance of differences between means. Table 2 presents the values of $t$ obtained for the various comparisons. As in the case of number of questions required, none of the differences between groups of two and of four is significant. Differences between individuals and groups of two on the third and fourth days are significant at the .05 level; differences between individuals and groups of four on all except the first day are significant at the same level or beyond.

A score for all four days was obtained for each individual or single group by taking the median time required for the 16 problems. The means of these scores were 5.06 for individuals, 3.70 for groups of two, and 3.15 for groups of four. The values of $t$ given in Table 2 show that the difference between the first and second mean is significant at the .05 level, and between the first and third mean at the .01 level.

Group performance was superior to individual performance in terms of elapsed time to solution. However, if, instead, an analysis is made in terms of number of man-minutes required for solution, the nature of the results obtained changes sharply. The number of man-minutes for a problem will, of course, be equal to the elapsed time multiplied by the number of persons in the group. In terms of man-minutes, the mean of the scores for all four days was 5.06 for individuals, 7.40 for groups of two, and 12.60 for groups of four. Since the variances for these three means were clearly not homogeneous, the use of $t$ was not appropriate for testing the significance of the obtained differences. Instead, $t'$ was employed (1). Both the difference between individuals and groups of two and the difference between groups of two and groups of four are significant at the .02 level. The difference between individuals and groups of four is significant at the .001 level. Clearly, in terms of man-minutes, the performance of individuals was superior to that of groups of two or of four; in addition, the performance of groups of two was superior to that of groups of four.

A supplementary question of some interest is whether the member of a group of two or of four getting the correct answer asked significantly more questions than the other member or members of the group. An analysis for all four days combined showed that for groups of two, the individual getting the correct answer asked an average of 1.55 questions more than the individual who failed to get the answer. A $t$ of 5.04 with 14 $df$ shows this to be significantly

different from zero at the .001 level. However, it may be plausibly argued that in making this comparison, the final question which identified the correct object should be excluded. Before asking it, the individual had correctly formulated the answer. If the final answer is excluded, the difference is reduced from 1.55 to .55. This yields a $t$ of 1.74 and is not significantly different from zero.

A similar analysis was done for groups of four. When the final question is included, the mean difference between the number of questions asked by the individual getting the answer and the average number asked by the other three members was 1.53. With a $t$ of 6.50, this is significantly different from zero at the .001 level. Excluding the final question reduces the mean difference to .53. However, with a $t$ of 2.25, this is still significantly different from zero at the .05 level. There appears to be some tendency for the member of a group of four getting the correct answer to ask more questions, even excluding the final question, than do other members of the group.

*Individual versus group practice.*— The third question which the experiment was intended to answer was whether improvement in individual performance occurs more rapidly with individual practice or with practice as a member of a group. To answer this question, all *S*s worked alone on the fifth day. As before, the score for each individual was the median number of questions required to solve the four problems. The mean of these scores for the 15 *S*s who had previously worked alone was 20.8;[3] for the 30 who had worked in pairs, 19.3; and for the 60 who had been members of groups of four, 19.1. None of the differences among these means is significant. Nor were any of the differences significant among the corresponding means on the fifth day for number of failures or for time scores. Learning went on as well in groups of two or of four as in individual practice.

## Discussion

The results obtained show that there is rapid learning of the skill involved in the game. The question now arises as to just what it is that is learned. To determine this, a qualitative analysis of the kinds of questions asked on successive days will be necessary. In a second experiment, now in progress, a complete record of all questions asked is being made in order that such an analysis can be carried out.

Group performances were superior to individual performance in terms of number of questions, number of failures, and elapsed time per problem; but the performance of groups of four was not superior to that of groups of two, except in terms of the number of failures to reach solution. Whether one could confidently have predicted such group superiority is questionable: Individual members of the group might have failed to make effective use of the information yielded by questions asked by other members; if this had been the case, the number of questions required by a group would have been larger, rather than smaller, than that required by an individual.

The fact that there were negligible differences between groups of two and of four either in number of questions

[3] Comparison of this mean for the fifth day with that for the fourth day (20.8 versus 18.1) shown in Fig. 1 may raise the question: Why should the performance on the fifth day be inferior to that on the fourth day in view of the fact that the conditions under which these 15 individuals worked were the same on both days? However, the difference between these two means is not significant ($t = 1.04$).

or in elapsed time strongly suggests that the optimum size group is not larger than four. Proof of this will require further experimentation with other size groups. Additional experiments are also needed to determine whether the optimum size group is similar for other types of problems.

The question may be raised as to why there was a significant difference between groups of two and of four in number of failures to reach solution, this in spite of the fact that there were negligible differences in number of questions or elapsed time. A possible explanation is that increasing the number of participants from two to four reduces the probability of a persisting wrong set resulting in complete failure. For an individual, a wrong set once established may make it impossible to solve the problem. The probability that a wrong set would be established simultaneously for all participants would be smaller for a group of four than for a group of two.

Although group performances were superior to individual performance in terms of elapsed time to solution, the performance of individuals was superior to that of either size group in terms of number of man-minutes required for solution. The practical implications of this fact should not be overlooked. It appears probable that there are many kinds of problems which a group will solve more quickly than an individual. If elapsed time in hours, weeks, or months is the primary consideration, then such problems should be undertaken by groups. However, it appears equally probable that few of those same problems will be solved more efficiently in terms of man-minutes or man-hours by groups than by individuals. If a group of two is to solve a problem more efficiently than an individual in these latter terms, it must solve it in less than half the elapsed time required by the individual. Similarly, a group of four to be more efficient must solve the problem in less than one-fourth the elapsed time required by the individual. The importance of this point appears to be frequently overlooked.

What it is that accounts for the superiority of group as compared to individual performance in terms of number of questions or elapsed time remains to be determined. The suggestion may be made that the superiority of the group is due to the performance of the best member of the group. If one were to pick the most able individual from each of 15 groups of four, it would be expected that the performance of these 15 individuals would be superior to that of 15 individuals chosen by random sampling. The mean number of questions required by groups of four on the fourth day was 13.6. The mean of the best individual performances on the fifth day by former members of each of the 15 groups of four was 14.8, not significantly different from 13.6. This fact would seem to support the suggestion just made. However, this comparison is not fully valid. Which former member of a group of four had the best performance on the fifth day very probably depended partly on ability and to a considerable extent on chance. Selecting the best individual performance from each of the 15 groups thus capitalizes on chance in a way that reduces the mean obtained; it may yet be true that the mean performance of the 15 groups would be superior to that of the best individuals in each of the 15 groups.

That the superior performance of the group is not simply a function of the performance of the best member of the group is suggested by another consideration. If this were the case, then the larger the group, the better on

the average should be the performance of the best member on the basis of sampling alone; hence the larger the group, the better should be the performance. The negligible differences obtained between groups of two and of four fail to confirm this expectation.

It may be expected that other factors such as broader range of relevant information, greater flexibility in approach, etc., are at least partly responsible for the superiority of group over individual performance. What these factors are and how they operate to produce an optimum size for a group can be determined only by additional experimentation.

An interesting supplementary question is whether the member of a group who obtains the right answer does so largely because he asks more questions than the other members of the group. The data obtained show that the number of questions asked by the member of a group of two obtaining the correct answer does not differ significantly from the number asked by the other member. A difference significant at the .05 level was found between the number asked by the member of a group of four obtaining the correct answer and the mean number asked by the other three members. However, this significant difference was only a matter of .53 questions per problem. It seems doubtful that getting the right answer is primarily due to the asking of more questions either in groups of two or of four.

The results obtained on the fifth day showed that learning resulting in improvement in individual performance occurred as rapidly with individual practice as with practice as a member of a group of two or of four. This fact, of course, should not be taken to mean that improvement is qualitatively the same under the different conditions. It may or may not be.

## Summary and Conclusions

The game of "Twenty Questions" was employed in an experiment on problem solving. A total of 105 *Ss* were assigned by chance to solve such problems working either alone, in pairs, or in groups of four. There were 15 individual *Ss*, 15 groups of two, and 15 groups of four. Each individual or group was given four problems a day for four successive days. On the fifth day, all *Ss* worked alone, each being given four problems.

Both the number of questions and the time required to solve each problem were recorded. Problems not solved in 30 questions were counted as failures.

1. In terms of number of questions, rapid improvement occurred in the performance both of individuals and of groups. By the fourth day, the curves appeared to be flattening out. Similar results were obtained in terms both of number of failures and of time per problem.

2. Group performances were superior to individual performance in terms of number of questions, number of failures, and elapsed time per problem; but the performance of groups of four was not superior to that of groups of two, except in terms of the number of failures to reach solution.

3. In terms of man-minutes required for solution, the performance of individuals was superior to that of groups; the performance of groups of two was superior to that of groups of four.

4. Improvement in individual performance occurred as rapidly with individual practice as with practice as a member of a group.

(Received for priority publication August 18, 1952)

## References

1. Cochran, W. G., & Cox, G. M. *Experimental designs.* New York: Wiley, 1950.
2. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika,* 1946, **11,** 97–105.
3. Lindley, E. H. A study of puzzles with special reference to the psychology of mental adaptation. *Amer. J. Psychol.,* 1897, **8,** 431–493.
4. Shaw, M. E. A comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.,* 1932, **54,** 491–504.

# JOURNAL OF EXPERIMENTAL PSYCHOLOGY

## The issues listed below are available for sale

| Year | Volume | Available Numbers | | | | | | Price per Number | Price per Volume |
|---|---|---|---|---|---|---|---|---|---|
| 1916 | 1 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1917 | 2 | 1 | 2 | – | 4 | 5 | – | $1.25 | $5.00 |
| 1920 | 3 | – | 2 | – | 4 | 5 | – | $1.25 | $3.75 |
| 1921 | 4 | 1 | – | 3 | 4 | 5 | 6 | $1.25 | $6.25 |
| 1922 | 5 | – | – | 3 | 4 | 5 | – | $1.25 | $3.75 |
| 1923 | 6 | – | – | – | – | – | – | —— | —— |
| 1924 | 7 | 1 | 2 | – | – | 5 | – | $1.25 | $3.75 |
| 1925 | 8 | 1 | 2 | 3 | 4 | – | 6 | $1.25 | $6.25 |
| 1926 | 9 | – | 2 | 3 | – | 5 | 6 | $1.25 | $5.00 |
| 1927 | 10 | – | – | 3 | 4 | 5 | 6 | $1.25 | $5.00 |
| 1928 | 11 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1929 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1930 | 13 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1931 | 14 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1932 | 15 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1933 | 16 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1934 | 17 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1935 | 18 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1936 | 19 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1937 | 20 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 21 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1938 | 22 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 23 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1939 | 24 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 25 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1940 | 26 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 27 | 1 | 2 | 3 | 4 | 5 | – | $1.25 | $6.25 |
| 1941 | 28 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 29 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1942 | 30 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 31 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1943 | 32 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 33 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1944 | 34 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1945 | 35 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1946 | 36 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1947 | 37 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1948 | 38 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1949 | 39 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1950 | 40 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1951 | 41 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 42 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1952 | 43 | By subscription, $14.00 (2 vols.) | | | | | | $1.25 | —— |
| | 44 | | | | | | | | |
| | | List price, Volumes 1 through 42 | | | | | | | $269.00 |
| | | 30% discount | | | | | | | 80.70 |
| | | Net price, Volumes 1 through 42 | | | | | | | $188.30 |

**Information about the Journal of Experimental Psychology:** as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

**Information about prices:** when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

*Address orders and subscriptions to*

## AMERICAN PSYCHOLOGICAL ASSOCIATION

333 Sixteenth Street N.W. Washington 5, D. C.